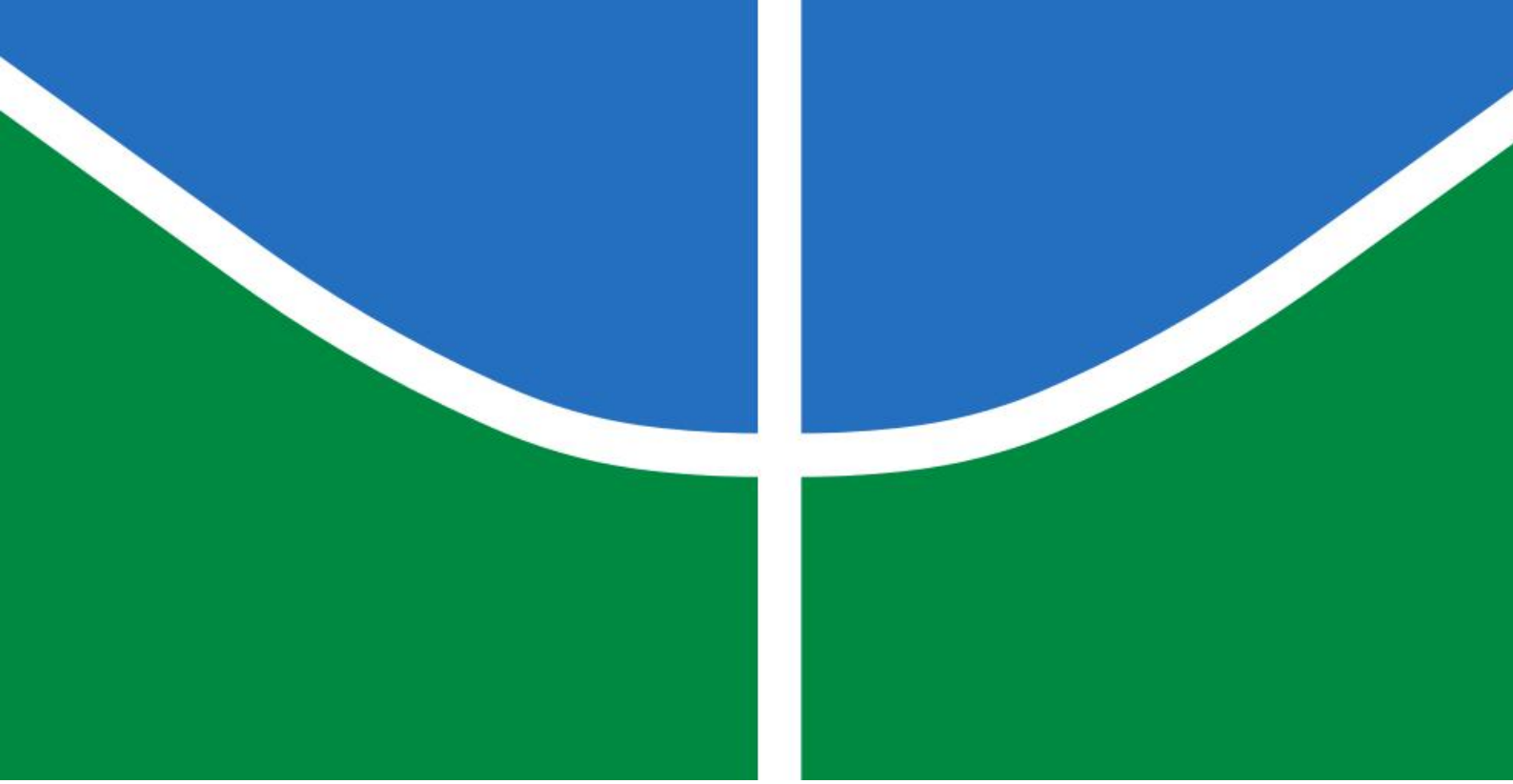

PROTOCOLO DE AUTENTICAÇÃO CONTÍNUA MULTIMODAL COM USO DE ELETROCARDIOGRAFIA PARA AMBIENTES DE COMPUTAÇÃO MÓVEL EM NUVEM

SILAS LEITE ALBUQUERQUE

TESE DE DOUTORADO EM ENGENHARIA ELÉTRICA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

PROTOCOLO DE AUTENTICAÇÃO CONTÍNUA MULTIMODAL COM USO DE ELETROCARDIOGRAFIA PARA AMBIENTES DE COMPUTAÇÃO MÓVEL EM NUVEM

SILAS LEITE ALBUQUERQUE

ORIENTADOR: PAULO ROBERTO DE LIRA GONDIM, DR.

COORIENTADOR: CRISTIANO JACQUES MIOSSO, DR.

TESE DE DOUTORADO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGEE.TD-164/20

BRASÍLIA/DF: ABRIL - 2020

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

PROTOCOLO DE AUTENTICAÇÃO CONTÍNUA MULTIMODAL COM USO DE ELETROCARDIOGRAFIA PARA AMBIENTES DE COMPUTAÇÃO MÓVEL EM NUVEM

SILAS LEITE ALBUQUERQUE

TESE DE DOUTORADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR.

APROVADA POR:

_____

Prof. Paulo Roberto de Lira Gondim - D.C. - ENE/FT/UnB
Orientador

_____

Prof. Adson Ferreira da Rocha, PhD - ENE/FT/UnB
Membro Interno

_____

Prof. Joel José Puga Coelho Rodrigues, PhD - UFPI
Membro Externo

_____

Renato Mariz de Moraes, PhD - CIN/UFPE
Membro Externo

BRASÍLIA, 06 DE MARÇO DE 2020.

FICHA CATALOGRÁFICA

ALBUQUERQUE, SILAS LEITE

Protocolo de autenticação contínua multimodal com uso de eletrocardiografia para ambientes de computação móvel em nuvem [Distrito Federal] 2020.

xxvi, 233p., 210 x 297 mm (ENE/FT/UnB, Doutor, Tese de Doutorado - Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

1.Segurança da Informação 2.Computação móvel em nuvem

3.Autenticação 4.Biometria

I. ENE/FT/UnB II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

ALBUQUERQUE, S. L. (2020). Protocolo de autenticação contínua multimodal com uso de eletrocardiografia para ambientes de computação móvel em nuvem. Tese de Doutorado em Engenharia Elétrica, Publicação PPGEE.TD-164/20, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 233p.

CESSÃO DE DIREITOS

AUTOR: Silas Leite Albuquerque.

TÍTULO: Protocolo de autenticação contínua multimodal com uso de eletrocardiografia para ambientes de computação móvel em nuvem.

GRAU: Doutor ANO: 2020

_____

Silas Leite Albuquerque
SQN 303, Bloco B, Asa Norte.
70.735-020 Brasília - DF - Brasil.

Dedicado aos meus pais (*in memoriam*), por sempre terem incentivado, em mim, o gosto pelo estudo.

AGRADECIMENTOS

Agradeço ...

... à minha família, pelo apoio constante;

... ao meu orientador, pelas várias orientações e pela infinita paciência;

... ao meu coorientador, pelos direcionamentos preciosos;

... aos demais professores do Departamento de Engenharia Elétrica, pelas dicas importantes.

RESUMO

A Computação Móvel em Nuvem (MCC, do inglês *Mobile Cloud Computing*), é uma especialização da Computação em Nuvem (CC, do inglês *Cloud Computing*) na qual, de maneira geral, os usuários acessam a nuvem por meio de equipamentos móveis. Considerando a grande complexidade da MCC, os seus problemas de segurança da informação apresentam muitas facetas e, particularmente a área de autenticação das diversas entidades envolvidas é algo que tem sido bastante discutido nos últimos anos e apresenta uma vasta gama de possibilidades.

Um dos grandes desafios existentes nesse contexto está relacionado aos processos de autenticação que devem ocorrer entre usuários e provedores de serviços de nuvem (CSP, do inglês *Cloud Service Providers*). Caso esses processos sejam inadequados, certamente surgirão muitos problemas que vão desde danos financeiros a usuários ou provedores de serviços de comércio móvel (do inglês *Mobile Commerce* ou *M-Commerce*) até a morte de pacientes que dependem de serviços de saúde móvel (*M-Health*, do inglês *Mobile Healthcare* ou *M-Healthcare*).

A criação de processos confiáveis de autenticação que minimizem os problemas indicados passa por questões como: fazer uso de vários fatores de autenticação (autenticação multifatorial ou multimodo) para aumentar a eficácia do processo; utilizar técnicas não intrusivas de autenticação e manter usuários autenticados continuamente perante os provedores de serviços de MCC. Focalizando as questões apontadas, a biometria provê recursos capazes de apoiar as necessidades de autenticações multifatoriais, não intrusivas e contínuas em diversos cenários.

Este trabalho, após explorar algumas bases conceituais, apresenta uma proposta de protocolo de autenticação contínua multifatorial com uso de eletrocardiografia para ambientes de MCC, a qual é combinada com uma senha e a Identidade Internacional do Assinante Móvel (IMSI, do inglês *International Mobile Subscriber Identity*).

Em termos de autenticação biométrica, o protocolo é baseado no monitoramento de informações fisiológicas interpretadas a partir de eletrocardiogramas (ECG) de uma base de dados com 108 usuários. Para a classificação dos ciclos cardíacos reconhecidos nesses ECG, são utilizadas e comparadas técnicas de aprendizado de máquina (ML, do inglês

Machine Learning) baseadas em Máquinas de Vetores de Suporte (SVM, do inglês *Support Vector Machines*), Busca pelo Vizinho mais Próximo (NNS, do inglês *Nearest Neighbor Search*), *Adaptative Boost* (*ou AdaBoost*) e *Robust Boost*.

Para a comparação das técnicas de ML aplicadas sobre a eletrocardiografia é adotada uma técnica de subamostragem aleatória que considera quatro métricas de análise: acurácia, precisão, sensibilidade e *F1-score*. Os resultados experimentais permitem concluir sobre um melhor desempenho da SVM em termos de acurácia (94,7%), sensibilidade (97,1%) e *F1-score* (94,9%); por outro lado, a NNS, apesar de não apresentar resultados tão bons quanto a SVM nestas métricas, apresenta uma melhor precisão (99,5%). Em comparação ao estado da arte observado em alguns trabalhos acadêmicos, os resultados indicados situam-se entre os três melhores.

Os resultados obtidos com o método proposto e em particular com os classificadores aplicados às características extraídas de sinais de ECG sugerem a viabilidade de autenticação contínua e não intrusiva baseada em ECG em sistemas multimodais. O ineditismo desta abordagem, em relação à literatura científica anterior, inclui a combinação da análise do ECG com o uso do IMSI vinculado ao chip do usuário de equipamentos móveis, bem como a uma senha fornecida pelo usuário no início da autenticação. Destaca-se ainda a avaliação de outros classificadores não analisados anteriormente para esta aplicação, bem como uso de uma forma melhorada de normalização das características temporais de ECG.

Como possíveis trabalhos futuros, cabe ressaltar a simulação sistemática do protocolo proposto em um simulador de redes, bem como a elaboração de novo algoritmo de detecção de pontos fiduciais que não focalize somente os ECG com ciclos que seguem o modelo indicado neste trabalho.

ABSTRACT

Mobile Cloud Computing (MCC), is a Cloud Computing (CC) specialization in which, in general, users access cloud through mobile equipment. Considering the great complexity of MCC, its information security problems have many facets and, particularly the authentication of the various entities involved is something that has been widely discussed in recent years and presents a wide range of possibilities.

One of the major challenges in this context is related to authentication processes that must occur between users and Cloud Service Providers (CSP). If these processes are inadequate, many problems will certainly occur, ranging from financial loss to users or providers of M-Commerce (Mobile Commerce) services to the death of patients who depend on M-Health (Mobile Healthcare) services.

Creation of reliable authentication processes that minimize the indicated problems involves issues such as: making use of various authentication factors (multifactorial or multimode authentication) to increase process effectiveness; use of non-intrusive authentication techniques and keep users continuously authenticated with MCC service providers. Focusing on these issues, biometrics provides resources capable of supporting the needs of multifactorial, non-intrusive and continuous authentication in different scenarios.

This work, after exploring some conceptual bases, presents a proposal for a multifactorial continuous authentication protocol using electrocardiography for MCC environments, which is combined with a password and the International Mobile Subscriber Identity (IMSI).

In terms of biometric authentication, the protocol is based on monitoring physiological information interpreted from electrocardiograms (ECG) obtained from a database with 108 users. For classification of cardiac cycles recognized in these ECG, Machine Learning (ML) techniques based on Support Vector Machines (SVM), Nearest Neighbor Search(NNS), Adaptative Boost (or AdaBoost) and Robust Boost are used and compared.

To validate ML techniques applied to the electrocardiography, a random sub sampling technique is adopted, which considers four metrics: accuracy, precision, sensitivity and F1-score. Experimental results allow us to conclude about a better performance of SVM in terms of accuracy (94.7%), sensitivity (97.1%) and F1-score (94.9%); on the other hand, NNS, despite not presenting as good results as the SVM in these metrics, presents a better precision (99.5%). In comparison to the state of the art observed in some academic works, the results indicated are between the three best.

The results we obtained using the proposed method and in particular the classifiers applied to the ECG features indicate the viability of continuous and noninvasive authentication based on ECG, in multimodal systems. The innovative aspect of this approach, with respect to the previous scientific literature, includes the combination of ECG analysis with the use of the IMSI in the mobile user's chip, as well as to the password the user provides in the beginning of the authentication. We also emphasize the evaluation of classifiers that were not previously considered in the literature for this kind of application, as well as an improved normalization of the ECG temporal features.

As possible future work, we suggest the systematic evaluation of the proposed protocol in a network simulator, as well as the implementation of a new fiducial points algorithm that can operate on ECG signals with patters differing from the standard ones we indicated in this work.

SUMÁRIO

LISTA DE TABELAS XVI

LISTA DE FIGURAS XVIII

LISTA DE ALGORITMOS XXI

LISTA DE TRECHOS DE CÓDIGO XXI

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIAÇÕES XXII

LISTA DE VARIÁVEIS, CONSTANTES E FUNÇÕES XXIV

1- INTRODUÇÃO 1

1.1- CONTEXTUALIZAÇÃO E MOTIVAÇÃO 1

1.2- O PROBLEMA E AS QUESTÕES DE PESQUISA 5

1.3- OBJETIVOS 6

1.4- PRINCIPAIS CONTRIBUIÇÕES 6

1.5- ARGUMENTO DA TESE 7

1.6- ORGANIZAÇÃO 7

2- FUNDAMENTOS DE COMPUTAÇÃO EM NUVEM E SEGURANÇA DA INFORMAÇÃO 9

2.1- CONSIDERAÇÕES GERAIS 9

2.2- MODELOS DE COMPUTAÇÃO EM NUVEM 10

2.2.1- Modelo baseado em serviços 10

2.2.2- Modelo baseado nas formas de implantação 11

2.3- COMUNICAÇÕES NO CONTEXTO DA COMPUTAÇÃO EM NUVEM 12
2.4- COMPUTAÇÃO MÓVEL EM NUVEM 12
2.4.1- Definição 12
2.4.2- Vantagens da MCC 15
2.4.3- Aplicações da MCC 16
2.4.4- Modelos de Aplicações para MCC 19
2.5- SEGURANÇA DA INFORMAÇÃO E OS AMBIENTES DE CC 22
2.5.1- Confiança computacional 22
2.5.2- Pilares da segurança da informação 32
2.5.3- Outros aspectos de segurança envolvendo MCC 36
2.6- CONCLUSÃO 37
3- FUNDAMENTOS DE AUTENTICAÇÃO E BIOMETRIA 39
3.1- PROCESSOS CONVENCIONAIS DE AUTENTICAÇÃO 39
3.1.1- Usuário / Senha 41
3.1.2- Número de Identificação Pessoal 42
3.1.3- Token 43
3.1.4- Certificado Digital 44
3.2- ASPECTOS DE BIOMETRIA 45
3.2.1- Aspectos Gerais 46
3.2.2- Tipos de biometria 50
3.3- AUTENTICAÇÃO BASEADA EM BIOMETRIA 56

3.3.1- Autenticação baseada em impressão digital 57

3.3.2- Autenticação baseada em reconhecimento facial.....58

3.3.3- Autenticação baseada na verificação da voz.....60

3.3.4- Autenticação baseada na verificação de sinais cardíacos.....61

3.4- ASPECTOS DE AUTENTICAÇÃO CONTÍNUA.....61

3.4.1- Autenticações contínua X estática e intrusiva X não intrusiva.....61

3.4.2- Níveis de autenticação contínua.....63

3.5- AUTENTICAÇÃO MULTIMODAL 65

3.6- CONCLUSÃO.....66

4- FUNDAMENTOS DE AUTENTICAÇÃO BASEADA EM ELETROCARDIOGRAFIA E DE APRENDIZADO DE MÁQUINA 67

4.1- ASPECTOS DO FUNCIONAMENTO DO CORAÇÃO.....67

4.1.1- Anatomia do coração.....67

4.1.2- O sistema de condução do coração 71

4.1.3- Eletrocardiografia.....72

4.2- BIOMETRIA BASEADA EM ELETROCARDIOGRAFIA 75

4.2.1- Aspectos gerais 75

4.2.2- Sensores cardíacos.....76

4.3- APRENDIZADO DE MÁQUINA (*MACHINE LEARNING*) 77

4.3.1- Fundamentos.....77

4.3.2- Máquinas de Vetores de Suporte 81

4.3.3- *Boosting* 86

4.3.4- Procura pelo vizinho mais próximo 89

4.4- CONCLUSÃO 90

5- PROTOCOLOS DE AUTENTICAÇÃO - TRABALHOS RELACIONADOS 91

5.1- PROPOSTAS DE ESQUEMAS, PROTOCOLOS OU MODELOS DE AUTENTICAÇÃO PARA AMBIENTES DE CC E MCC 91

5.1.1- Esquema de autenticação eficiente para serviços de MCC 91

5.1.2- Esquema melhorado para autenticação eficiente de serviços de MCC distribuídos 93

5.1.3- Serviço de autenticação para CC baseado em padrões de digitação 94

5.1.4- Sistema eficiente de autenticação para MCC baseado em vários fatores 96

5.1.5- Comparação entre os trabalhos apresentados 98

5.2- PROPOSTAS DE UTILIZAÇÃO DE ELETROCARDIOGRAFIA PARA IDENTIFICAÇÃO E AUTENTICAÇÃO 99

5.2.1- Reconhecimento biométrico de ECG utilizando uma abordagem baseada em SVM 99

5.2.2- Autenticação para equipamentos móveis baseada em ECG 101

5.2.3- Utilização de fluxos de eletrocardiogramas em tempo real para a autenticação contínua 104

5.2.4- Autenticação bimodal baseada em ECG e impressão digital 106

5.2.5- Comparação entre os trabalhos apresentados 107

5.3- CONCLUSÃO 109

6- PROPOSTA DE PROTOCOLO DE AUTENTICAÇÃO MULTIMODAL CONTÍNUA 110

6.1- CONSIDERAÇÕES INICIAIS 110

6.2- DEFINIÇÃO DE MODELO DE REDE 112

6.3- DEFINIÇÃO DO MODELO DE AUTENTICAÇÃO.....114

6.3.1- Registro.....115

6.3.2- Autenticação 117

6.3.3- Atualização 121

6.4- ASPECTOS ESPECÍFICOS DO PROCESSO BASEADO EM ELETROCARDIOGRAFIA.....124

6.4.1- Método Extrair *features* de ECG 125

6.4.2- Método Treinar classificador.....132

6.4.3- Método Autenticar.....133

6.5- CONCLUSÃO.....135

7- AVALIAÇÃO DA PROPOSTA E RESULTADOS.....136

7.1- VALIDAÇÃO DO PROCESSO DE AUTENTICAÇÃO BASEADO EM ELETROCARDIOGRAFIA UTILIZANDO BASE DE ECG PÚBLICA 136

7.1.1- Bases de dados de Eletrocardiogramas.....136

7.1.2- Preparação das *features* utilizadas no modelo de autenticação 141

7.1.3- Processo de subamostragem aleatória para validação da autenticação 143

7.1.4- Análise estatística dos resultados em termos de eficácia do processo 164

7.2- CONCLUSÃO.....166

8- CONCLUSÕES E RECOMENDAÇÕES 167

8.1- CONCLUSÕES GERAIS 168

8.2- PRINCIPAIS CONTRIBUIÇÕES.....171

8.3- RECOMENDAÇÕES PARA TRABALHOS FUTUROS 172

REFERÊNCIAS BIBLIOGRÁFICAS.....173

APÊNDICES 186

APÊNDICE A - TABELAS COMPLETAS DOS VALORES DAS MÉTRICAS DE ANÁLISE REFERENTES AOS CLASSIFICADORES.....187

APÊNDICE B - ARTIGO SUBMETIDO A PERIÓDICO COM JCR 207

LISTA DE TABELAS

Tabela 4.1 - Variáveis utilizadas neste capítulo 68

Tabela 5.1 - Comparação entre os trabalhos gerais analisados 99

Tabela 5.2 - Comparação entre os trabalhos analisados 108

Tabela 6.1 - Variáveis utilizadas neste capítulo 111

Tabela 7.1 - Variáveis e constantes utilizadas neste capítulo.....137

Tabela 7.2 - Características da base de ECG utilizada neste trabalho 138

Tabela 7.3 - Números de ciclos reconhecidos por usuário da base (ordenado por $NC_{u)}$ ...140

Tabela 7.4 - Dados consolidados sobre as quantidades de ciclos reconhecidos 141

Tabela 7.5 - *SVM* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise 149

Tabela 7.6 - *AdaBoost* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise 150

Tabela 7.7 - *Robust Boost* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise 151

Tabela 7.8 - *NNS* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise 152

Tabela 7.9 - Mapa de calor representando os valores das métricas de análise e os desvios padrão em todos os classificadores utilizados 154

Tabela 7.10 - Mapa de calor análogo ao da Tabela 7.9 considerando apenas os casos em que $NCu = NCMax$.....155

Tabela 7.11 - SVM - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.....156

Tabela 7.12 - *AdaBoost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos. 157

Tabela 7.13 - *Robust Boost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos. 158

Tabela 7.14 - NNS - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.....159

Tabela 7.15 - Valores-p dos testes de normalidades de cada métrica em cada classificador considerando todos os 108 usuários.....164

Tabela 7.16 - Valores-p dos testes de comparação entre os classificadores. 165

LISTA DE FIGURAS

Figura 1.1 - Arquitetura típica dos serviços de MCC 3

Figura 2.1 - Arquitetura de computação em nuvem orientada a serviços 10

Figura 2.2 - Arquitetura típica de MCC 13

Figura 2.3 - Processo de Offloading para a nuvem.....14

Figura 2.4 - exemplo de modelo de confiança para transações eletrônicas 28

Figura 3.1 - Classificação de métodos de autenticação 40

Figura 3.2 – FAR X FRR 48

Figura 3.3 - Impressões digitais.....50

Figura 3.4 – medição de parâmetros da face 51

Figura 3.5 – Procedimento de reconhecimento da geometria da mão 52

Figura 3.6 – Iris e Retina.....53

Figura 3.7 – Dispositivos móveis com leitores de impressão digital.....57

Figura 3.8 –Modelos de equipamentos móveis com câmeras 59

Figura 4.1 –Posição e forma do coração 69

Figura 4.2 –Câmaras e válvulas do coração 70

Figura 4.3 – Representação de ECG de dois indivíduos diferentes.....73

Figura 4.4 – Alinhamento de ondas de ECG de um mesmo indivíduo.....74

Figura 4.5 –Modelos de sensores cardíacos comerciais.....76

Figura 4.6 - Exemplo de classificação binária.....81

Figura 4.7 - Hiperplanos de uma SVM treinada para uma classificação binária 83

Figura 4.8 - Margens variáveis entre hiperplanos de uma SVM. 84

Figura 4.9 - Superfície de separação entre classes em um exemplo não linearmente separável 84

Figura 5.1- Modelo de rede 92

Figura 5.2 - Arquitetura do sistema de autenticação baseado em *keystrokes* 95

Figura 5.3 - Arquitetura do sistema 98

Figura 5.4 - *Features* extraídas dos pontos fiduciais (à esquerda) e descritores morfológicos (à direita) 101

Figura 5.5 - Pontos fiduciais detectados pelo algoritmo (à esquerda) e *features* calculadas (à direita) 103

Figura 5.6 - Estrutura geral do sistema de autenticação baseada em ECG 105

Figura 5.7 - Algoritmo de autenticação bimodal 107

Figura 6.1 - Modelo de rede considerado 113

Figura 6.2 - Fase de registro 117

Figura 6.3 - Subfase de autenticação inicial 118

Figura 6.4 - iteração da subfase de autenticação contínua 120

Figura 6.5 - Fase de atualização por meio de ambiente controlado 122

Figura 6.6 - Fase de atualização durante sessão autenticada 123

Figura 6.7 - Ciclo cardíaco típico 125

Figura 6.8 - Extração de *features* de ECG 126

Figura 6.9 - Efeito da aplicação do filtro "Butterworth" 127

Figura 6.10 - Sinal filtrado de ECG com picos e vales detectados.....128

Figura 6.11 - Distâncias consideradas para o cálculo das *features* de ECG 129

Figura 6.12 - Amostra de sinal de ECG de 5 s 130

Figura 6.13. Diferenças dos comprimentos dos ciclos.....131

Figura 7.1 - Números de ciclos reconhecidos por usuário da base (ordenado por NC_u)....139

Figura 7.2 - *SVM* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise 149

Figura 7.3 - *AdaBoost* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise 150

Figura 7.4 - *Robust Boost* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise 151

Figura 7.5 - *NNS* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise 152

Figura 7.6 - Tempos médios gastos pelas várias amostragens aleatórias para o treinamento e para a classificação 163

LISTA DE ALGORITMOS

Algoritmo 4.1 - Procedimento geral da técnica de *boosting* 87

Algoritmo 6.1 - Funcionamento geral do protocolo de autenticação proposto 114

Algoritmo 6.4 - Autenticação baseada em NNS.....134

Algoritmo 8.1 - Passos da subamostragem aleatória 144

LISTA DE TRECHOS DE CÓDIGO

Trecho de código 6.1 - Utilização do filtro *Butterworth* no Matlab 126

Trecho de código 6.2 - Utilização da função *findpeaks* 127

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIAÇÕES

ADB	Base de dados de autenticação (do inglês, *Authentication Database*)
bpm	Batimentos por minuto
CAC	Centro de autenticação contínua (do inglês, *Continuos Authentication Center)*
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	Computação em núvem (do inglês, *Cloud computing*)
CSP	Provedor de serviços de nuvem (do inglês, *Cloud Service Provider*)
DSM	Mineração de fluxo de dados (do inglês, *Data Stream Mining*)
ECG	Eletrocardiograma
EER	Taxa de iguadade de erros (do inglês, *Equal Error Rate*)
FAR	Taxa de aceitação falsa (do inglês, *False Acceptance Rate*)
FC	Frequência cardíaca
FRR	Taxa de rejeição falsa (do inglês, *False Rejection Rate*)
HS	Sensor cardíaco (do inglês, *Heart Sensor*)
IaaS	Infraestrutura como serviço (do inglês, *Infrastruture as a Service*)
ICP	Infraestruturas de Chaves Públicas
IMEI	Número de identidade internacional do equipamento móvel (do inglês, *International Mobile Equipment Identiy*)
IMSI	Número de identidade internacional do assinante móvel (do inglês, *International Mobile Subscriber Identiy*)
kNN	Vizinho de proximidade de ordem k (do inglês, *k-Nearest Neighbor*)
MC	Computação móvel (do inglês, *Mobile Computing*)
MCC	Computação Móvel em Nuvem (do inglês, *Mobile Cloud Computing*)
MD	Equipamento móvel (do inglês, *Mobile device*)
ME	Equipamento móvel (do inglês, *Mobile equipment*)
ML	Aprendizado de máquina (do inglês, *Machine Learning*)
NNS	Busca pelo vizinho mais próximo (do inglês, *Nearest Neighbor Search*)
PAA	Autenticação ciente de privacidade (do inglês, *Privacy-Aware Authentication*)
PaaS	Plataforma como serviço (do inglês, *Platform as a Service*)
PKI	Infraestruturas de Chaves Públicas (do inglês, *Public Key Infraestructures*)
PR	Reconhecimento de padrões (do inglês, *Pattern Recognition*)
QRS	Onda do ciclo cardíaco compreendida entre os pontos fiduciais Q e S
RBF	Função de base radial (do inglês, Radial Basis Function)
RC	Centro de registro (do inglês, *Registration Center*)
RL	Aprendizado por reforço (do inglês, *Reinforcement Learning*)
SaaS	*Software* como serviço (do inglês, *Software as a Service*)
SCG	Gerador de cartão esperto (do inglês, *Smart Card Generator*)

SIM	Módulo de identidade do assinante (do inglês, *Subscriber Identity Module)*
SL	Aprendizado supervisionado (do inglês, *Supervised learning*)
SNA	Sistema Nervoso Autônomo
SP	Provedor de serviços (do inglês, *Sevice Provider*)
SSL/TLS	Camada de Soquete Segura / Segurança da Camada de Transporte (do inglês, *Secure Socket Layer* / *Transport Layer Security*)
SSO	Assinatura única (do inglês, *Single Sign On*)
SVC	Classificação por vetores de suporte (do inglês, *support vector classification*)
SVM	Máquina de vetores de suporte (do inglês, *Suport Vector Machine*)
TAR	Taxa de aceitação verdadeira (do inglês, *True Acceptance Rate*)
TTP	Terceira parte confiável (do inglês, *Trusted Third Party*)
UE	Equipamento do usuário (do inglês, *user equipment*)
UL	Aprendizado não supervisionado (do inglês, *Unsupervised learning*)
WPA	Acesso *Wi-Fi* protegido (do inglês, *Wi-Fi Protected Access*)

LISTA DE VARIÁVEIS, CONSTANTES E FUNÇÕES

Variável/ constante/ função	Descrição
Capítulo 4	
tp	Número de verdadeiros positivos
tn	Número de verdadeiros negativos
fp	Número de falsos positivos
fn	Número de falsos negativos
Acc	Valor da acurácia
Pre	Valor da precisão
Sen	Valor da sensibilidade
$F1s$	Valor do escore F1
l	Rótulo de um classificador
$x_1, x_2,..., x_m$	Variáveis que representam pontos discretos de um domínio de análise de um problema de classificação e que também podem ser entendidas como exemplos a serem submetidos em um treinamento de classificador
$C_0, C_1,..., C_m$	Classes possíveis de uma classificação também chamadas de rótulos
$y_1,y_2,..., y_m$	Variáveis que assumem os valores das Classes possíveis de uma classificação
$sign( )$	Função sinal que fornece o sinal do resultado de $( )$ ou 0 caso o resultado seja nulo
w	Vetor perpendicular ao hiperplano ótimo de separação de classes em uma SVC em um espaço n-dimensional
$h_1, h_0, h_{-1}, h_1', h_0', h_{-1}'$	Hiperplanos de um espaço n-dimensional
w	Vetor perpendicular ao hiperplano ótimo no espaço de n dimensões
b	Valor escalar ou constante
K	Função de *kernel*
ϕ	função que transforma um vetor x de um espaço n-dimensional em um vetor equivalente em espaços de dimensões mais altas (até infinitas)
x'_j	Variáveis não rotuladas (cuja classe correspondente é desconhecida)
$\boldsymbol{x}_i^T$	Vetor de variáveis de treinamento em um domínio transformado
$\gamma, r \ e \ d$	Parâmetros do *kernel* de uma SVM
D	Distribuição amostral considerada no treinamento de um classificador
$Class$	Classificador específico
lo_{exp}	Função de perda exponencial
f	Função que retorna o rótulo de x
$E_{\sim D}[f( )]$	Esperança matemática de uma função f
$I( )$	Função que cujo resultado é 1 se $( )$ é verdadeiro e 0 caso contrário
T	Número total de iterações de um treinamento ou quantidade de classificadores fracos que fazem parte de um processo de *boosting*
t	Uma iteração de um treinamento ou índice que um dos classificadores que fazem parte de um processo de *boosting*
α	Peso dado ao classificador que está sendo treinado

θ	Margem a ser atingida no processo de minimização de erro
M	Espaço *n*-dimensional
p	Ponto de M
S	Conjunto de pontos p
q	Ponto de M que serve de referência para os demais

Capítulo 6

Pwd	Senha do usuário
H( )	Aplicação de uma função de condensação sobre um conteúdo
IdAut	Identificador de autenticação da sessão de comunicação segura entre o CSP e o usuário
$E_0[t]$	Vetor de dados brutos de ECG obtido do HS e formado por valores de amplitudes do ECG que representam potenciais elétricos (em mV)
$E_1[t]$	Vetor de dados filtrados (atenuação de ruído) de ECG
IDX	Vetor que contem os índices de $E_1[t]$correspondentes aos picos e vales detectados
c	Índice do ciclo cardíaco considerado na análise
LP[c], P[c], Q[c], R[c], S[c], T[c] e TP[c]	Vetores contendo os pontos fiduciais detectados em cada ciclo *c*
RLP[c], RP[c],RQ[c], RS[c], RT[c] e RTP[c]	Vetores contendo as distâncias, em termos de índices, entre os pontos considerados do ciclo *c*
RSA[c] e RQA[c]	Vetores contendo as distâncias, em termos de amplitudes, entre os pontos considerados do ciclo *c*
$f_1[c], f_2[c],.. , f_8[c]$	Vetores de *Features* extraídas para cada ciclo *c*
NC	Número de ciclos considerados para o treinamento do classificador
μ_{fj}	Média da "j-ésima" *feature* considerando todos os ciclos do treinamento
Tol_j	Tolerância referente à "j-ésima" *feature* para uma classificação com NNS
Acertos	Quantidade de acertos na comparação entre *features* e médias armazenadas
AcertosMin	Quantidade mínima de acertos adequada para o processo

Capítulo 7

Nu_{Tot}	Total útil de usuários da base de dados de ECG, ou seja, número de usuários constantes da base que forneceram os ECG utilizados neste trabalho.
u	Índice de um usuário considerado da base
$NC_{u\,Max}$	Número máximo de ciclos utilizados de cada usuário
NC_u	Número de ciclos cardíacos reconhecidos por usuário
NC_{Tot}	Número total de ciclos reconhecidos considerando Nu_{Tot}
NC_{Max}	Número máximo de ciclos que podem ser adotados considerando Nu_{Tot} e $NC_{u\,Max}$
i	Índice de um ciclo cardíaco de um usuário
$c_{i,u}$	"i-ésimo" ciclo cardíaco de um usuário

Nf_{Tot}	Número total de *features*, por ciclo, utilizadas neste trabalho
j	Índice de uma *feature* do $c_{i,u}$
$f_{j,i,u}$	"j-ésima" feature de um $c_{i,u}$
$\Delta t_{ECG\ Min}$	Duração mínima dos ECG considerados
M_u	Matriz de $f_{j,i,u}$ por usuário
M	Matriz de $f_{j,i,u}$ de todos os usuários
V	Matriz coluna de todos *u* relacionada à *M*
$Conj_{Class}$	Conjunto formado pelos 4 classificadores utilizados neste trabalho
$Class$	Classificador utilizado em uma análise específica
$Conj_{Metr}$	Conjunto formado pelas 4 métricas de análise utilizadas neste trabalho
$Metr$	Métrica utilizada em uma análise específica
Ns_{Tot}	Número total de sessões de validação da técnica de subamostragem aleatória, por usuário
s	Índice da sessão de validação da técnica de subamostragem aleatória
$SAM_{u,s}$	Subamostra aleatória da matriz *M* referente ao usuário *u* e à sessão *s*
k	Índice (que também representa o tamanho) do grupo de sessões de validação da técnica de subamostragem aleatória
$Gs_{k,u}$	"k-ésimo" grupo de sessões de validação da técnica de subamostragem aleatória aplicada sobre o usuário u

1- INTRODUÇÃO

A autenticação de usuários em ambientes de computação móvel em nuvem (do inglês, *Mobile Cloud Computing* - MCC) representa um desafio para cuja superação diversos protocolos têm sido empregados. Dentre tais protocolos, destacam-se os que se baseiam em autenticação multifatorial, por permitirem a verificação de diferentes fatores associados a um usuário que pretenda utilizar serviços de um dado provedor.

Este trabalho apresenta a proposta e a avaliação de um protocolo que se baseia no emprego de fatores convencionais (senha e identificação do assinante móvel), de forma combinada com sinais biométricos baseados em eletrocardiogramas. Tomando por base uma população de 108 usuários, técnicas de processamento de sinais e aprendizado de máquina são empregadas, permitindo obter um protocolo que, além de atuar de forma integrada sobre uma arquitetura de MCC, realiza suas operações de forma não intrusiva e contínua, favorecendo, dessa forma, a melhoria da segurança da informação.

Merece destaque o fato de temas tratados nesta tese fazerem parte de sugestões de "trabalhos futuros" encontradas em pesquisas recentes publicadas em periódicos de relevância ([FERRAG, 2019], por exemplo).

1.1- CONTEXTUALIZAÇÃO E MOTIVAÇÃO

No mundo extremamente interconectado de hoje, a computação em nuvem é uma realidade irrefutável, uma tendência mundial e tem se desenvolvido rapidamente como uma das tecnologias de rede mais poderosas [HE, 2018]. Considerando a simples possibilidade de armazenamento remoto de informações ou a complexa estrutura que viabiliza plataformas inteiras de provimento de serviços conforme as necessidades específicas de cada usuário, essa tecnologia surgiu, em grande parte, para permitir que os equipamentos diretamente utilizados pelos indivíduos não necessitem mais possuir grandes capacidades de armazenamento nem de processamento, pois essas capacidades passaram a ser disponibilizadas pela nuvem.

Essa nuvem é dinâmica, flexível e configurável, o que permite aos usuários a obtenção de serviços por demanda a partir de uma série de recursos de rede compartilhados e que podem ser alocados com esforço mínimo [MOHSIN, 2017].

Por outro lado, a computação móvel (do inglês, *Mobile Computing* - MC) remonta à ideia da possibilidade de transportar facilmente um computador enquanto o mesmo pode ser utilizando dentro de suas potencialidades de processamento, armazenamento e comunicação.

Os equipamentos computacionais móveis (ou somente equipamentos móveis, do inglês, *Mobile Equipment* - ME) tendem a possuir hardware mais modesto, menor, mais leve e, por conseguinte, na maioria das vezes, com menos recursos computacionais que os equipamentos fixos, são excelentes exemplos de potenciais interfaces para o acesso à computação em nuvem, além de também poderem servir como sensores remotos para determinados tipos de processos vinculados aos seus usuários (captura de dados de contexto, sinais vitais, voz, imagem, etc.).

Da junção da flexibilidade da computação móvel com as capacidades de armazenamento e processamento da computação em nuvem, surgiu a chamada Computação Móvel em Nuvem. Nela, a maior parte do processamento e do armazenamento dos dados dos equipamentos móveis é transferida para plataformas centralizadas de computação localizadas na nuvem, o que permite que esses equipamentos com menores capacidades computacionais, desde que conectados à nuvem por meio das diversas tecnologias disponíveis (redes locais sem fio, do inglês, *Wireless Local Area Networks* - WLAN, redes móveis celulares 3G, 4G, 5G, etc.), possam executar aplicações mais complexas e acessar novos recursos e serviços [JIANG, 2016], pois passam a funcionar, apenas, como interfaces dos usuários com capacidades adequadas de comunicação.

Uma arquitetura típica e simplificada de MCC e dos serviços que podem ser fornecidos é apresentada na Figura 1.1. Nela são representados os usuários móveis com seus equipamentos acessando os serviços de nuvem computacional (infraestrutura como serviço, do inglês, *Infrastruture as a Service* - IaaS; plataforma como serviço, do inglês, *Platform as a Service* - PaaS; software como serviço, do inglês, *Software as a Service* - SaaS) por meio de conexões de acesso convencional à Internet.

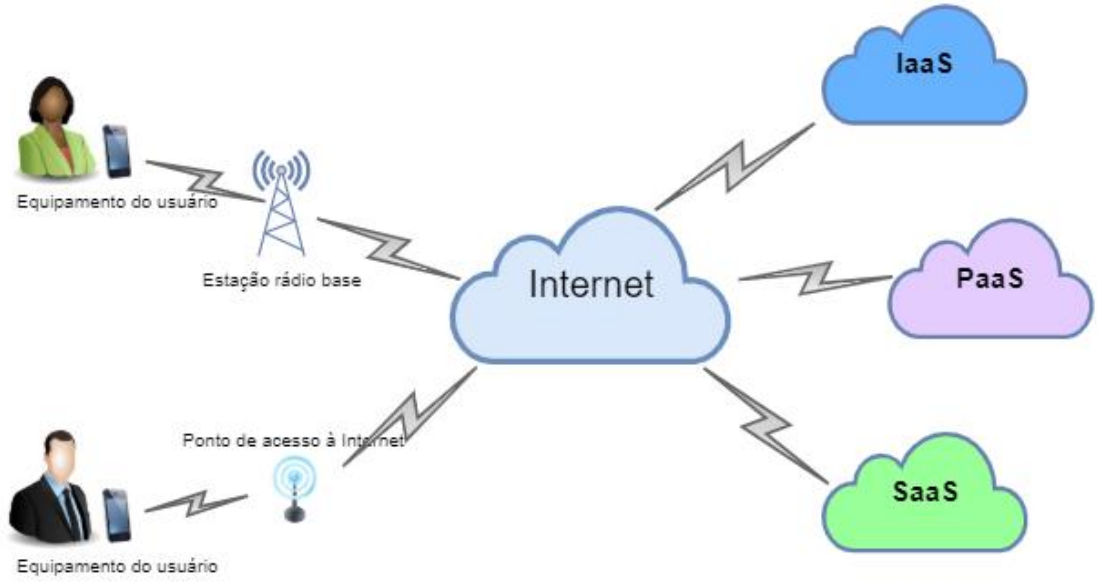

Figura 1.1 - Arquitetura típica dos serviços de MCC

Esse estilo de utilização dos ambientes computacionais tem evoluído nos últimos anos em muito pela larga expansão do mercado de telefones inteligentes (do inglês, *smartphones*), o que fez aumentar tremendamente a quantidade de usuários da computação móvel e trouxe, vinculada, a necessidade de melhoria, tanto qualitativa quanto quantitativa, das infraestruturas voltadas para esse segmento.

Há vários exemplos de aplicação prática da computação móvel em nuvem (por exemplo, comércio móvel, do inglês, *M-Commerce*; aprendizado móvel, do inglês, *M-Learning*; jogo móvel, do inglês, *M-Gaming*; saúde móvel, do inglês, *M-Health* ou *M-Healthcare*; etc.). Um dos mais interessantes é a *M-Health*, que diz respeito à utilização da MCC voltada para a manutenção da saúde dos seus usuários. As redes de sensores médicos sem fio (do inglês, *Wireless Medical Sensor Networks* – WMSN) [HAYAJNEH, 2016], por exemplo, são soluções que podem ser integradas às arquiteturas de MCC para viabilizar o controle remoto feito por profissionais de saúde sobre pacientes que se encontram, muitas vezes, distantes geograficamente de unidades de atendimento, conferindo rapidez e comodidade para pessoas que necessitam de cuidados médicos, mas que não podem ou não desejam permanecer em ambientes voltados para esse fim.

Os sensores sem fio citados, além de poderem ser utilizados para atender a objetivos voltados para a saúde, também têm aplicação em outras áreas, e uma que particularmente interessa ao tema explorado nesse trabalho é a autenticação biométrica.

Apesar dos benefícios da computação em nuvem, tanto no contexto convencional quanto no contexto móvel, problemas que outrora não existiam passaram, nos últimos anos, a ser também uma realidade. A área de segurança da informação, por exemplo, é algo que apresenta uma série de brechas que não existiam em outras arquiteturas. Pelo fato da computação em nuvem estar baseada na utilização remota de informações, quer seja para simplesmente armazená-las quer seja para processá-las, vários pilares fundamentais da segurança da informação passaram a ser ameaçados: a confidencialidade, a integridade, a disponibilidade e a autenticidade, por exemplo.

Considerando o grande aumento, em comparação às arquiteturas isoladas, da quantidade de elementos que existem entre a interface do usuário e o local onde as informações serão armazenadas e/ou processadas na nuvem, um dos assuntos que tem merecido muita atenção nos meios acadêmico e comercial é a autenticação das partes integrantes de uma arquitetura de MCC.

Nos últimos anos, muitos trabalhos têm tido o objetivo de analisar e/ou propor soluções de autenticação para ambientes de CC e/ou MCC ([MOHSIN, 2017], [JANNATI, 2017], [DEY, 2016], [AUDITHAN, 2017], [AL-RUBAIE, 2016] e [ABDULWAHID, 2015], por exemplo). Esses trabalhos indicam a existência de muitos métodos e protocolos de autenticação largamente utilizados, reconhecidos globalmente por muitos pesquisadores e até padronizados por alguns órgãos normativos. Dentre esses métodos, algumas técnicas de biometria têm se mostrado alternativas bastante interessantes, não só pelo fato de dependerem exclusivamente de aspectos intrínsecos do ser humano a ser autenticado, mas também pelo fato de permitirem, além de uma autenticação em um momento inicial de uma sessão ou transação (acesso de um UE à infraestrutura da MCC, por exemplo), a possibilidade de continuidade dessa autenticação (o usuário pode permanecer sendo autenticado durante todo o período da sessão) de forma transparente ou não intrusiva para o indivíduo.

A biometria baseada em eletrocardiografia é uma dessas opções. As características utilizadas nesse tipo de técnica biométrica são únicas, podem ser medidas, ocorrem sem

que haja a intervenção voluntária de qualquer pessoa (nem do próprio indivíduo de onde são extraídas) e são difíceis de serem falsificadas [FALCONI, 2016], [REZGUI, 2016], [CAMARA, 2017]. Dessa forma, essa biometria apresenta muitos aspectos desejáveis para um bom método a ser empregado em atividades de autenticação de pessoas perante sistemas.

Apesar de a eletrocardiografia ser uma boa opção para a autenticação, a interpretação equivocada dos dados derivados dos ECG pode gerar distorções bastante inadequadas. Para minimizar esse problema, destaca-se o aprendizado de máquina (do inglês, *Machine Learning* - ML) como alternativa flexível e robusta a partir da qual é possível fornecer boas soluções para questões complexas de classificação e obter bons resultados de eficácia e eficiência para a autenticação.

1.2- O PROBLEMA E AS QUESTÕES DE PESQUISA

Antes de definir o problema alvo deste trabalho, foram definidas algumas questões de pesquisa:

- O emprego da eletrocardiografia permite autenticação de forma eficaz e eficiente?
- É possível utilizar alguma técnica de aprendizado de máquina como parte de um processo de autenticação baseado em eletrocardiografia?
- Caso seja, quão eficazes são essas técnicas de ML?
- É possível projetar um protocolo de autenticação forte, multimodal, contínua e não intrusiva que se baseie em eletrocardiografia?
- De que forma a área de segurança da informação aplicada à computação móvel em nuvem pode ser beneficiada pelo emprego da eletrocardiografia?

Considerando os diversos aspectos analisados e as diversas ideias surgidas durante os estudos e discussões, chegou-se à conclusão que o trabalho deve visar à resolução do seguinte problema:

- Autenticar, eficiente e eficazmente, usuários e provedores de serviços de MCC, uns perante os outros, de maneira pouco intrusiva e contínua, utilizando, de maneira conjugada, métodos de processamento de sinais, autenticação e aprendizado de máquina, aplicados à eletrocardiografia para garantir uma autenticação forte.

1.3- OBJETIVOS

Focalizando o problema apresentado, o presente trabalho pretende atingir os seguintes objetivos:

- Realizar uma ampla revisão bibliográfica sobre os temas computação em nuvem, computação móvel em nuvem, fundamentos de segurança da informação e suas aplicações nos ambientes de CC e MCC, fundamentos de autenticação, biometria, eletrocardiografia e aprendizado de máquina;
- Como principal objetivo e contribuição, propor um novo protocolo de autenticação forte, mútua e contínua para ambientes de MCC;
- Avaliar a eficácia e a eficiência de partes do protocolo proposto;
- Realizar uma análise comparativa dessa proposta com outras encontradas na literatura.

1.4- PRINCIPAIS CONTRIBUIÇÕES

Em linhas gerais, pode-se considerar como contribuições da presente tese:

- A proposta de um protocolo de autenticação contínua forte combinando três fatores que não foram combinados em outros trabalhos;
- A proposta de um processo de atualização dos padrões de ECG armazenados na base de autenticação;
- A implementação de um novo algoritmo de detecção de pontos fiduciais;
- A utilização de nova técnica de normalização para a obtenção de parâmetros de ECG para os classificadores;
- A utilização de *ensembles* para a classificação de sinais de eletrocardiografia;
- A aplicação da técnica de subamostragem aleatória para a validação do processo de autenticação baseado em eletrocardiografia;
- A utilização de várias métricas de análise (acurácia, precisão, sensibilidade e *f1-score*) para a avaliação dos processos de classificação empregados.

Essas contribuições são detalhadas de maneira mais completa no capítulo de conclusão desta tese.

Além das contribuições citadas, ao longo das atividades realizadas durante o doutorado, foram aceitos e publicados os trabalhos indicados em [ALBUQUERQUE, 2010], [ALBUQUERQUE, 2012] e [ALBUQUERQUE, 2016] (este, especificamente, foi inserido no estrato A2 da classificação de periódicos "Qualis" da CAPES, em 2016).

1.5- ARGUMENTO DA TESE

De maneira geral, o argumento da tese (do inglês, *thesis statement*) explorado neste trabalho pode ser definido como:

- É possível propor um novo protocolo de autenticação contínua, multimodal (com fatores independentes) e transparente (não intrusiva) para usuários de serviços de computação móvel em nuvem porque ainda há opções pouco exploradas de combinações de técnicas de autenticação (particularmente de biometria) e algoritmos de aprendizado de máquina que viabilizam níveis de acurácia, precisão e sensibilidade adequados.

1.6- ORGANIZAÇÃO

Além de iniciar pela presente introdução, este trabalho está assim organizado:

- O capítulo 2 apresenta temas referentes à computação em nuvem, à computação móvel em nuvem, a fundamentos de segurança da informação e suas aplicações em ambientes de CC;
- O capítulo 3 descreve alguns fundamentos da autenticação e de técnicas de biometria, além de explorar a aplicação da biometria para a autenticação e descrever alguns conceitos de autenticação contínua e multimodal;
- O capítulo 4 apresenta aspectos do funcionamento do coração e da captura de sinais elétricos oriundos deste órgão, fundamentos da biometria baseada em sinais de eletrocardiografia e, por último, conceitos e técnicas de aprendizado de máquina;
- O capítulo 5 descreve brevemente e compara alguns trabalhos relacionados aos temas explorados nesta tese;
- O capítulo 6 apresenta a proposta de protocolo de autenticação, detalhando as suas diversas fases, os métodos e algoritmos utilizados e alguns modelos empregados;
- O capítulo 7 apresenta a avaliação da proposta de protocolo focalizando,

particularmente, os aspectos vinculados à eletrocardiografia;

- O capítulo 8, finalmente, indica algumas conclusões e contribuições obtidas, além de sugerir, para o futuro, alguns trabalhos e pesquisas;
- O Apêndice A contem algumas tabelas e gráficos utilizados no capítulo de avaliação da proposta;
- O Apêndice B apresenta uma proposta de artigo submetida a um periódico com JCR.

2- FUNDAMENTOS DE COMPUTAÇÃO EM NUVEM E SEGURANÇA DA INFORMAÇÃO

2.1- CONSIDERAÇÕES GERAIS

A ideia de nuvem na computação remonta à representação gráfica de um ambiente sobre o qual não se tem muito conhecimento (não se sabe, com precisão, de que é formado) mas pelo qual passam informações com origens e destinos bem definidos. Essa representação é tipicamente utilizada quando o ambiente da nuvem não interessa muito para o tema que está sendo estudado ou sobre o qual está se dizendo algo.

Atualmente a computação em nuvem é uma tendência mundial por vários motivos, dentre os quais podemos destacar:

- Possibilidade de utilização remota de sistemas sem que estes estejam instalados no ambiente computacional diretamente utilizado pelo usuário;
- Transparência para o usuário em relação à infraestrutura necessária (*hardware*, sistema operacional, etc.) para o provimento de um determinado serviço;
- Possibilidade de utilização da computação por demanda, sem a necessidade de montagem de uma infraestrutura que poderá, no momento seguinte à execução do serviço, ficar ociosa;
- Delegação, para os provedores de serviços, de responsabilidades que, em uma arquitetura não baseada em nuvem, geram uma grande carga de trabalho para os usuários (indivíduos ou organizações) dos serviços e que os desviam de suas atividades finalísticas;
- Facilitação do compartilhamento de informações entre usuários dispersos geograficamente;
- Otimização de utilização dos recursos computacionais e demais recursos vinculados a esses (energia elétrica, espaço físico, etc.).

De uma forma geral, a computação em nuvem pode ser considerada a infraestrutura de computação da atual e das próximas gerações [DINH, 2013] e, muito além da ideia de um ambiente computacional de localização incerta, é interpretada como um tipo de

processamento baseado na web no qual recursos compartilhados, software e informações são fornecidos para os usuários sob demanda [FAN, 2011].

2.2- MODELOS DE COMPUTAÇÃO EM NUVEM

2.2.1- Modelo baseado em serviços

As vantagens citadas anteriormente devem-se, em muito, ao fato da computação em nuvem estar baseada na ideia de utilização dos recursos computacionais como serviços, não havendo, assim, a preocupação, do usuário final, com a montagem de infraestruturas, instalação de programas, alocação de recursos como memória e poder de processamento para a execução das tarefas necessárias ao atendimento das necessidades do usuário.

Nesse contexto, pode-se considerar que uma nuvem computacional seja um sistema de larga escala formado por redes de computadores distribuídos e baseado em servidores existentes em *datacenters* [DINH, 2013] e, de maneira geral, os serviços providos pela nuvem computacional podem ser classificados com base no conceito de camadas, conforme a arquitetura representada na Figura 2.1.

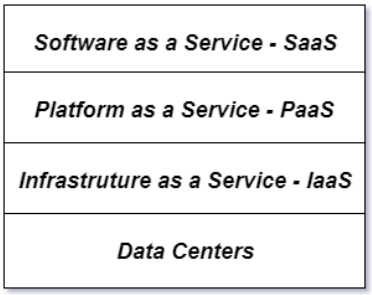

Figura 2.1 - Arquitetura de computação em nuvem orientada a serviços

- A camada de mais alto nível, Software como serviço - SaaS - representa o fornecimento de software para atender a determinados requisitos específicos. Nessa camada os usuários podem utilizar remotamente uma aplicação e acessar os dados a ela vinculados e pagar somente por essa utilização e por esse acesso.
- A camada da Plataforma como serviço - PaaS - pode oferecer, por exemplo, um ambiente integrado para o desenvolvimento, teste e implantação de aplicativos personalizados.

- A camada da Infraestrutura como serviço - IaaS - baseia-se no fornecimento de recursos computacionais (armazenamento de dados, poder de processamento e infraestrutura de redes) de tal forma que seu custo para um usuário que queira contratar esse tipo de serviço seja baseado somente nos recursos de fato utilizados. Um usuário pode, nesse contexto, contratar um determinado poder de processamento, uma capacidade de armazenamento e uma capacidade de transmissão de dados sem estar preocupado, de maneira direta, com a obtenção e gestão dos meios que de fato serão utilizados para o provimento desses recursos computacionais.
- A camada de *Data centers* representa as instalações físicas e toda a restante infraestrutura (servidores, *storages*, redes de comunicações, etc.) necessária para que a nuvem computacional funcione.

Cabe salientar, nesse contexto, que diversos dos provedores de serviços assinalados anteriormente valem-se de tecnologias de virtualização para viabilizar os recursos computacionais comentados, e essas tecnologias permitem uma maior flexibilidade no fornecimento dos serviços além de uma maior escalabilidade no contexto do atendimento das necessidades dos usuários. Com a utilização dessas tecnologias de virtualização e desde que a infraestrutura instalada suporte, hoje é possível de forma muito fácil e rápida aumentar, por exemplo, as capacidades de processamento e armazenamento necessárias a uma execução de uma determinada tarefa.

2.2.2- Modelo baseado nas formas de implantação

Outro modelo muito conhecido aplicado à computação em nuvem é o baseado nas formas de implantação da infraestrutura que suportará o fornecimento dos serviços citados no item anterior. Com base nesse modelo, as nuvens podem ser classificadas como [GOYAL, 2014]:

- Nuvem privada - este tipo representa as nuvens que estão sob a posse (ou foram alugadas) integral de uma só organização e, por esse motivo, todos os recursos dessa nuvem estão dedicados aos propósitos exclusivos da organização proprietária.
- Nuvem pública - o proprietário desse tipo de nuvem é um provedor de serviços que tem a finalidade de fornecer seus serviços para o público em geral, o que faz com

que toda a sua infraestrutura seja compartilhada por usuários de diversas origens diferentes.

- Nuvem comunitária - é um tipo semelhante à nuvem privada, entretanto os recursos da nuvem não são utilizados exclusivamente por uma organização, mas sim por uma comunidade fechada formada por entidades que têm interesses em comum.
- Nuvem híbrida - é uma combinação dos tipos citados anteriormente (de dois ou mais).

2.3- COMUNICAÇÕES NO CONTEXTO DA COMPUTAÇÃO EM NUVEM

Um aspecto muito importante que tem tornado a computação em nuvem algo cada vez mais real e útil é o aumento nas capacidades de transmissão da informação. Uma das ideias principais da CC baseia-se no fato dos equipamentos acessados pelos usuários, apesar de poderem ter baixos poderes computacionais (capacidade de processamento e armazenamento), deverem possuir conexões eficientes com a nuvem, haja vista que a informação necessita fluir da interface do usuário para a nuvem, e vice versa, da forma mais rápida possível para que seja armazenada e/ou processada ou para que o resultado do processamento seja devolvido para o interessado. Assim, quanto melhor for a rede de interconexão de dados existente entre a interface do usuário e a nuvem, maior será a probabilidade do serviço provido pela nuvem atender integralmente às expectativas do usuário.

2.4- COMPUTAÇÃO MÓVEL EM NUVEM

2.4.1- Definição

MCC é um paradigma para aplicações móveis onde o processamento e o armazenamento dos dados são transferidos para plataformas centralizadas de computação localizadas na nuvem [DINH, 2013], o que permite que equipamentos com menores capacidades computacionais (smartphones, por exemplo) possam executar aplicações mais complexas pois passam a funcionar, apenas, como interfaces dos usuários com capacidades adequadas de comunicação.

Nesse paradigma, dois conceitos bastante atuais da computação são misturados: a computação móvel e a computação em nuvem. Assim, boa parte das características dessas

duas áreas também é observada na computação móvel em nuvem. Os modelos de computação em nuvem citados anteriormente, por exemplo, aplicam-se integralmente à MCC com o acréscimo da ideia de que o terminal de utilização do usuário final que possui poucos recursos computacionais e que fará uso dos recursos da nuvem possui, na mobilidade, uma de suas principais características.

A Figura 2.2 ilustra uma arquitetura de MCC típica e mostra que os terminais móveis, para poderem acessar a infraestrutura convencional da nuvem computacional (*cloud controllers*, *data centers*, etc.) e provedores vinculados (*application service providers*, por exemplo), necessitam acessar a Internet por meio da estrutura provida por suas operadoras de rede (*Satellites*, *Access points*, *base transceiver stations*, *mobile network services*, etc.) e, uma vez tendo atingido a web, o acesso à CC será similar àquele realizado por outros tipos de terminais.

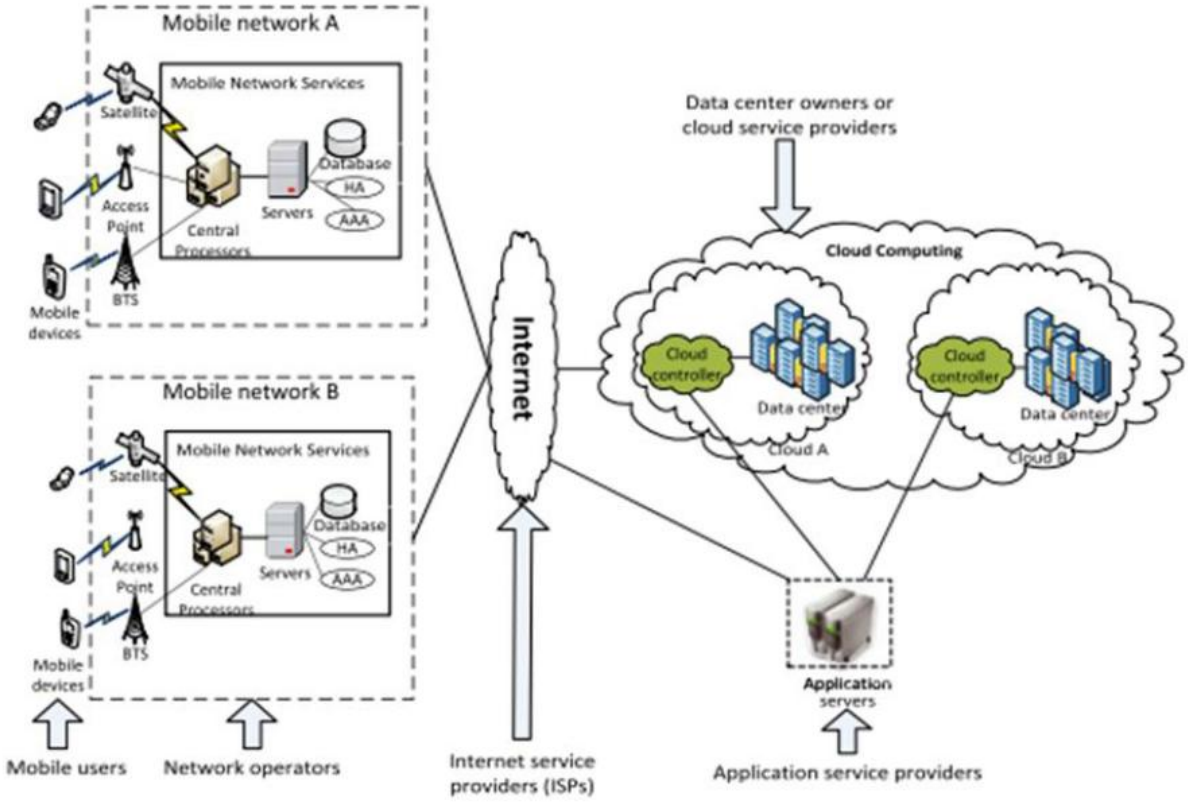

Figura 2.2 - Arquitetura típica de MCC
[DINH, 2013]

Observa-se que a maior parte tanto do processamento quanto do armazenamento das informações ocorre fora do equipamento móvel, na nuvem computacional.

De uma maneira geral, MCC [HRESTAK, 2014] pode ser descrito como o provimento de serviços de CC para um ambiente de equipamentos móveis, ou seja, o fornecimento das potencialidades de sistemas de armazenamento e processamento de dados distribuídos pelo mundo para equipamentos móveis.

Para que essas tarefas remotas possam ser realizadas, faz-se necessário um processo de escolha "do que, de como e de quando" as informações migrarão da interface para a nuvem. A esse processo dar-se o nome de *offloading* de aplicações para a CC [BAHL, 2012] e ele ocorre em um contexto no qual se tem, de um lado, um equipamento móvel, geralmente com capacidades limitadas e aplicações com grandes necessidades computacionais, e do outro, uma nuvem computacional, com capacidades quase ilimitadas e que pode ser alocada para atender às necessidades das aplicações do equipamento móvel.

A Figura 2.3 ilustra o processo no qual é tomada a decisão de se executar a aplicação no próprio equipamento móvel ou na nuvem, considerando as opções do *offloading* estar habilitado, de haver recursos disponíveis no móvel para a execução da aplicação e do processo de *offloading* ser favorável de uma forma geral.

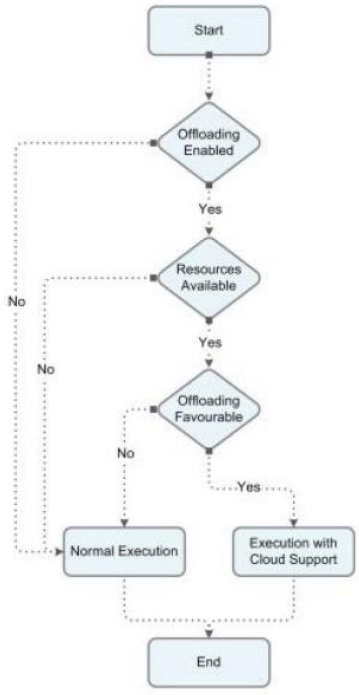

Figura 2.3 - Processo de Offloading para a nuvem
[KHAN, 2014]

2.4.2- Vantagens da MCC

Além de viabilizar a execução de certas aplicações que dependem de arquiteturas distribuídas e não poderiam ser executadas em um equipamento móvel isolado, a MCC apresenta uma série de outras vantagens (algumas específicas da MCC, outras relativas à CC de uma forma geral) como, por exemplo [SUO, 2013], [WAN, 2013]:

- Aumento do poder de processamento para a execução de aplicações no ambiente móvel - provavelmente o maior dos benefícios da utilização da MCC pois um equipamento móvel, considerando seus reduzidos tamanhos justificados pelo atendimento à necessidade de mobilidade, geralmente não possui grandes capacidades de processamento, e o *offloading* das tarefas que demandam muito processamento, do móvel para a nuvem, pode tornar viável o atendimento, a contento, de requisitos que não poderiam ser atendidos por equipamentos isolados e com pequeno poder de processamento;
- Aumento da capacidade de armazenamento utilizada pelos equipamentos móveis - análoga à vantagem anterior, a possibilidade de utilização da nuvem para armazenamento das informações é algo que torna praticamente ilimitadas as potencialidades de persistência de dados a serem utilizados pelos usuários dos equipamentos móveis; salienta-se que esta vantagem foi, se não a, uma das primeiras características a serem exploradas por meio da utilização de CC;
- Aumento do tempo de uso do equipamento móvel entre recargas - a partir do momento em que ocorre o *offloading* de tarefas para uma nuvem computacional, o móvel deixa de realizar, para atender aos requisitos das aplicações em execução, uma série de processos internos que consomem energia, o que permite uma maior duração do tempo no qual o móvel permanece ligado sem sofrer uma recarga em sua bateria;
- Diminuição da probabilidade de perda de dados - considerando que os dados dos usuários, no caso de utilização da nuvem computacional, são enviados para ambientes computacionais que estão sujeitos a processos de backup (processos esses que não são comumente realizados por usuários normais dos equipamentos móveis nos seus próprios ambientes), a probabilidade de que algum dado seja perdido é diminuída de forma relevante;

- Provisionamento dinâmico de recursos - é uma vantagem inerente às arquiteturas baseadas em nuvem que viabilizam recursos sob demanda para os usuários, quer sejam móveis ou fixos, e que permite a otimização da utilização compartilhada das diversas potencialidades disponíveis na nuvem além de uma tarifação proporcional mais justa e econômica;
- Aumento da escalabilidade - considerando a vantagem anterior, fica relativamente simples para um usuário da nuvem computacional aumentar ou diminuir a alocação dos recursos necessários para atender às demandas de suas aplicações;
- Otimização na utilização dos recursos da nuvem - essa vantagem é, em muito, viabilizada pelas arquiteturas de virtualização sobre as quais muitas nuvens computacionais se baseiam e que permitem a "alocação inteligente" de recursos por meio do fornecimento das capacidades exatas necessitadas pelos usuários para atenderem às suas aplicações (nesse caso, as capacidades que sobrarem serão repassadas para outros usuários ou, em não havendo demanda, serão colocadas em um regime de espera, do inglês, *stand by*);
- Facilidade no compartilhamento de informações - considerando que as informações, em um ambiente de MCC, migram em grande parte para a nuvem, o compartilhamento dessas fica facilitado a partir do momento em que não é mais necessário enviar uma informação de um usuário para o outro, mas tão somente disponibilizar um link para essa informação que já se encontra na nuvem.

2.4.3- Aplicações da MCC

Como foi possível perceber no item anterior, a junção da computação móvel com a computação em nuvem traz uma série de benefícios herdados de ambas as áreas [WAN, 2013], e esses benefícios viabilizam a execução de uma série de aplicações que, em um contexto no qual o equipamento móvel funcione de forma isolada, atenderiam de forma pífia (ou sequer atenderiam) aos requisitos dos usuários.

A seguir são listadas algumas aplicações para as quais a utilização da MCC traz uma série de benefícios [DINH, 2013], [WAN, 2013].

2.4.3.1- Comércio móvel (do inglês, *M-Commerce*)

Essa é uma modalidade de comércio eletrônico (do inglês, *E-commerce*) na qual o consumidor utiliza um equipamento móvel para realizar a sua parte da transação. Considerando a grande complexidade das operações envolvidas em uma transação comercial e a conseguinte necessidade de processamento por parte do equipamento do consumidor (móvel), a passagem de parcelas da aplicação utilizada para um ambiente de nuvem favorece de forma relevante a execução da transação comercial, quer seja no contexto da velocidade do processo quer seja no tocante à confiabilidade das operações efetuadas.

Salienta-se que a "terceirização" de funções que deveriam ser, em uma arquitetura de *M-Commerce* convencional, executadas diretamente no equipamento móvel, traz, dentre outros, problemas de segurança da informação, e esses problemas serão explorados mais a frente neste trabalho.

2.4.3.2- Aprendizado móvel (do inglês, *M-Learning*)

Análogo à aplicação do item anterior, o *M-Learning* é um tipo de aprendizado eletrônico (do inglês, *E-Learning*), também conhecido como ensino eletrônico à distância, no qual as interações da maioria dos usuários (alunos, tutores, professores e etc.) ocorre por meio de equipamentos móveis.

Este tipo de aplicação, tal qual o *M-Commerce*, também está sujeito às limitações dos equipamentos móveis para viabilizar as funcionalidade ligadas ao ensino. A utilização de recursos da nuvem computacional, como a larga capacidade de armazenamento e de processamento, podem viabilizar serviços com uma maior riqueza de informações e de apresentação, o que gerará um ensino de melhor qualidade para os usuários.

Além disso, a facilidade de compartilhamento de informações que o uso da nuvem permite faz com que os contatos entre professores, tutores e alunos tornem-se mais simples, eficientes e eficazes.

2.4.3.3- Assistência móvel à saúde (do inglês, *M-Heathcare* ou *M-Health*)

A *M-Health* baseia-se, de forma geral, no fornecimento, para o paciente, de informações referentes à sua saúde, e para os provedores de assistência à saúde, da possibilidade de acompanhamento mais aproximado do estado dos pacientes.

Nesse contexto, a utilização da MCC para a *M-Heathcare*, além de minimizar os problemas relacionados às mesmas limitações dos equipamentos móveis indicadas nos exemplos anteriores (baixa capacidade de armazenamento e de processamento), também pode viabilizar, considerando as facilidades de compartilhamento de informações da nuvem, o monitoramento remoto e constante de pacientes que necessitem de um acompanhamento contínuo por parte dos provedores de assistência à saúde. É possível, por exemplo, utilizar equipamentos móveis para monitorar sinais vitais de pacientes gerando dados que são enviados para a nuvem computacional e observados por uma série de profissionais de especialidades médicas diferentes.

2.4.3.4- Jogo móvel (do inglês, *M-Gaming*)

A utilização de jogos em equipamentos móveis é algo que demanda grandes capacidades de processamento (para a realização da renderização de gráficos, por exemplo) e de armazenamento que nem sempre estão disponíveis nos próprios equipamentos. A utilização das potencialidades da MCC, nesse contexto, é algo que minimiza essas deficiências. Em uma situação extrema, é possível, por exemplo, delegar praticamente todo o processamento e armazenamento das informações referentes a um determinado jogo diretamente para nuvem, ficando o móvel (ou os móveis, no caso de jogos com mais de um participante) restrito apenas à apresentação dos resultados.

2.4.3.5- Outras aplicações

As possibilidades de aproveitamento das vantagens citadas no item 2.2.2. são muitas. Pode-se utilizar MCC, além dos exemplos citados, para: processamento de voz, processamento de vídeo, compartilhamento de fotos e vídeos capturados por meio dos móveis, compartilhamento de informações geográficas e outras informações vinculadas a essas, acionamento de "máquinas de busca" em vários contextos, verificação remota de vírus, etc.

2.4.4- Modelos de Aplicações para MCC

Um modelo de aplicação para MCC tem diversos objetivos possíveis como, por exemplo [KHAN, 2014]:

- a execução de aplicações não suportadas pelo ambiente do equipamento móvel isolado;
- Otimização do desempenho de aplicações (diminuição do tempo de computação);
- Economia de energia do equipamento móvel;
- Combinações dos itens anteriores.

Por outro lado, o fato de haver um determinado objetivo prioritário, por vezes, prejudica o atingimento de outros objetivos. Um exemplo típico disso ocorre ao tentar-se economizar energia do equipamento, o que muitas vezes gera uma degradação no desempenho das aplicações.

Assim, um modelo de aplicação é algo que deve ser adotado de tal forma que favoreça um ou mais objetivos predefinidos. Nesse contexto, na maioria das vezes, aqueles modelos que viabilizam vários objetivos têm prevalência em relação aos demais, pois dão suporte a uma quantidade maior de aplicações.

Em [KHAN, 2014] são detalhados alguns modelos que serão brevemente explicados a seguir.

2.4.4.1- *CloneCloud*

É um modelo que tem como objetivo prioritário a melhoria do desempenho das aplicações do móvel e baseia-se na descarga (do inglês, *offloading*) de partes da execução da aplicação (tarefas) diretamente para a nuvem. Esse modelo não demanda qualquer conversão das aplicações que se encontram no móvel pois pressupõe a existência de um clone do móvel posicionado na nuvem. Assim, o processo de *offloading* ocorre por meio do envio de partes inalteradas da execução da aplicação do móvel para o seu clone localizado na nuvem. Tendo recebido as partes citadas, o clone realiza o processamento necessário e devolve os resultados para o móvel. Para que tudo ocorra sem que sejam geradas inconsistências nos resultados do processamento, deve ser mantida uma sincronização precisa entre o móvel e o seu clone.

2.4.4.2- Modelo de Zhang et al.

Tal qual o anterior, este modelo também prioriza a melhoria do desempenho das aplicações do móvel [ZHANG, 2010], entretanto não está baseado no envio de partes simples das aplicações para um clone na nuvem, mas sim na técnica das "aplicações elásticas" na qual uma única aplicação é particionada em vários componentes chamados *weblets* que podem ser definidos como unidades funcionais independentes de uma aplicação. As *weblets* consideram processamento, armazenamento e comunicação, e podem ser executadas de maneira independente tanto no móvel quanto na nuvem.

2.4.4.3- µCloud

Este modelo prioriza a economia da energia do equipamento móvel baseada na redução do processamento computacional das aplicações no móvel por meio do *offloading* para a nuvem. Apesar de também utilizar o princípio da divisão das aplicações em componentes, tal qual o modelo anterior, os componentes das aplicações, neste modelo, diferem das *weblets*, pois são heterogêneos e interdependentes. A execução das aplicações, neste modelo, é sequencial e um componente só pode realizar seu processamento após o recebimento dos resultados do processamento do componente anterior.

2.4.4.4- Modelo de Satyanarayanan et al.

É um modelo que não busca otimizar o desempenho das aplicações tampouco gerar economia de energia do móvel, mas sim viabilizar a execução de aplicações em ambientes que apresentam restrições de recursos por meio do *offloading* para a nuvem [SATYANARAYANAN, 2009]. Neste caso, pelo fato do móvel não apresentar recursos suficientes para a execução da aplicação integralmente, apenas componentes da aplicação que exijam poucos recursos serão executados no móvel e o restante, que demanda recursos mais complexos, será executado no ambiente de nuvem. O modelo usa um conceito de máquina virtual, localizada em uma nuvem com recursos computacionais adequados e chamada de *cloudlet*, que recebe tarefas de processamento "pesado" vindas do equipamento móvel, as processa e devolve os resultados para o móvel.

2.4.4.5- Modelo de Giurgiu et al.

Como o anterior, este também é um modelo que se propõe a viabilizar a execução de aplicações em ambientes com poucos recursos computacionais por meio do *offloading* parcial para a nuvem [GIURGIU, 2009]. Este modelo baseia-se na técnica das "camadas distribuídas" na qual camadas funcionais são distribuídas entre o móvel (geralmente executa as partes da aplicação localizadas na camada de apresentação) e a nuvem (executa a aplicação no contexto das demais camadas) com o objetivo de diminuir a latência (tempo decorrido entre o início e o fim do *offloading*) e os atrasos causados pela transferência dos dados.

2.4.4.6- MAUI

Este é um modelo com múltiplos objetivos, sendo que o principal é viabilizar a economia de energia no móvel, entretanto também existe a preocupação em melhorar o desempenho da aplicação de uma forma geral. Para atingir esses objetivos, o modelo baseia-se no *offloading*, do móvel para a nuvem, de métodos que são menores e mais simples que os componentes e as unidades funcionais dos modelos anteriormente indicados. A transferência desses métodos desonera o processamento do móvel e gera atrasos menores que nos casos apontados anteriormente, o que permite, de um lado, a economia de energia do móvel (menor processamento), e do outro, uma melhoria no desempenho geral das aplicações (menores atrasos nos processos de *offloading*).

2.4.4.7- *ThinkAir*

Também é um modelo com múltiplos objetivos e assemelha-se, em parte, ao *CloneCloud*, sendo que, além de viabilizar o *offloading* de métodos do móvel para a nuvem a fim de diminuir o consumo de energia no primeiro (menor processamento), utiliza múltiplos clones na nuvem para paralelizar a execução da aplicação e, com isso, melhorar o desempenho como um todo.

2.4.4.8- Cuckoo

Semelhante aos dois anteriores, o *Cuckoo* também é um modelo com múltiplos objetivos (visa a diminuir o consumo de energia do móvel e melhorar o desempenho geral) e funciona especificamente sobre o sistema operacional *Android*. Sua arquitetura está

baseada na divisão da aplicação em serviços (demandam um processamento intensivo) e atividades (responsáveis pela interação móvel-nuvem) que podem ser executados tanto local quanto remotamente, dependendo da decisão do Gerenciador de Recursos do *Cuckoo* (do inglês, *Cuckoo Resource Manager* - CRM) após uma análise detalhada do contexto da execução da aplicação e da utilização dos recursos disponíveis.

2.5- SEGURANÇA DA INFORMAÇÃO E OS AMBIENTES DE CC

A garantia da segurança da informação no contexto da computação em nuvem, quer no modelo tradicional quer no móvel, é um dos aspectos mais complexos e de difícil solução que existem relacionados ao tema. Há vários problemas, alguns de mais simples solução, outros que ainda demandam muita pesquisa para que se chegue a resultados efetivos [ZISSIS, 2012]. A seguir serão exploradas algumas áreas relacionadas ao tema.

2.5.1- Confiança computacional

Para que seja possível falar de modelos de confiança a serem aplicados em ambientes de computação em nuvem, é importante explorar algumas definições que permitem um melhor entendimento do contexto geral.

Confiança

A primeira e mais relevante é a de confiança. A definição clássica encontrada no dicionário Houaiss da Língua Portuguesa é:

> *"...crença na probidade moral, na sinceridade afetiva, nas qualidades profissionais, etc., de outrem, que torna incompatível imaginar um deslize, uma traição, uma demonstração de incompetência de sua parte; crédito, fé..."*

Nesse sentido, percebe-se a nítida vinculação da confiança à crença ou à fé em aspectos morais de entidades em que se confia (do inglês, *trusted party*), o que torna inimaginável algum tipo de traição por parte dessas entidades em relação a quem confia (do inglês, *trusting party*). Apesar de esse conceito ser coerente com a vida cotidiana, ele remete a situações ligadas a relacionamentos humanos ("aspectos morais"), o que não combina diretamente com o que está sendo explorado neste trabalho, afinal as entidades envolvidas em ambientes de computação em nuvem não são somente pessoas, mas sim,

sistemas complexos formados também por hardware e software, além das próprias pessoas. Dessa forma, outras definições mais abrangentes devem ser consideradas.

Em [JOSANG, 2007] confiança é tratada considerando dois aspectos distintos e complementares: a esperança (uma entidade tem esperança que outra entidade aja em seu favor) e a dependência (uma entidade está disposta a depender de outra para atingir algum objetivo). Os dois enfoques são muito próximos e em ambos os casos, quanto maior a probabilidade, maior será a confiança.

Em [WANG, 2007] considera-se que confiança seja algo subjetivo e esteja vinculado à opinião das pessoas. Por esse motivo, é possível dizer que ela é passada de uma entidade para outra (transitividade) – "eu confio em você, você confia nele, logo eu confio nele" – o que pode configurar uma teia de confiança (do inglês, *web of trust*) [DATTA, 2003].

De uma forma geral, o conceito de confiança que interessa para esse trabalho deve considerar: a crença criada de forma direta (acordos formais) ou indireta (*web of trust*) em entidades com as quais deseja-se fazer alguma transação; a esperança que não haja algum tipo de traição da *trusted party*; e a dependência a qual a *trusting party* está disposta a se submeter.

Reputação

Outro conceito tão importante quanto o de confiança é o de reputação, que pode ser entendida como o que geralmente é dito em uma comunidade sobre uma entidade (deriva do comportamento perante outras entidades) [JOSANG, 2007] ou a opinião pública que existe acerca do caráter e do comportamento honesto de uma entidade em um determinado meio [WANG, 2007].

Confiança e reputação são conceitos complementares e é possível dizer que uma forma bastante válida de obter confiança em uma entidade é analisar sua reputação, da mesma forma, a boa reputação de uma entidade pode ser construída a partir da confiança que muitos depositam nesta entidade. Entretanto, apesar de parecerem proporcionais, confiança e reputação nem sempre são. Vejamos, por exemplo, as seguintes afirmações:

- “Confio em você por causa da sua boa reputação”;
- “Confio em você, independente da sua má reputação”;
- “Não confio em você, independente da sua boa reputação”.

Na primeira afirmação, a confiança deriva diretamente da reputação, entretanto nas duas seguintes, ela deriva de outra coisa que provavelmente é mais relevante que a própria reputação (um conjunto de leis e normas, por exemplo, que previna danos causados a uma entidade - que confia - por outra entidade - em quem se confia – pode ser suficiente para alguém confiar em outrem, independente de sua reputação).

Modelo de confiança

Um terceiro conceito importante no contexto da computação em nuvem é o de modelo de confiança, que pode ser entendido como uma estrutura formada por três componentes: as entidades, seus comportamentos e seus relacionamentos. As entidades representam as diversas partes envolvidas nos processos e se caracterizam principalmente por dois papeis: quem é acreditado (do inglês, *trustee*) - *trusted party* - e quem acredita (do inglês, *trustor*) - *trusting party*. Entretanto, dependendo da forma de relacionamento entre as partes, algumas entidades intermediárias com papeis específicos podem surgir para viabilizar, por exemplo, mecanismos de segurança. Vinculados às entidades estão os seus comportamentos, que podem gerar boa (ou má) reputação e ensejar maior ou menor confiança. Por último, os relacionamentos representam as diversas ligações existentes entre as entidades e viabilizam, além da própria avaliação dos comportamentos, o fornecimento de informações e a prestação de serviços de uma forma geral.

Cada componente deve estar adequado para viabilizar confiabilidade em níveis coerentes às transações que ocorrerão sobre esta estrutura. Essa adequação deve ser feita a partir de perguntas como [KOIEN, 2003]:

- Em que entidades deve-se confiar?
- De que forma deve-se confiar nessas entidades?
- Que tipos de características de segurança são necessárias para justificar confiança?

Essas questões podem ser respondidas de diversas formas, dependendo da situação analisada, e indicarão as premissas básicas que devem ser atendidas por um modelo de confiança que esteja em conformidade com o problema explorado.

Confiança em ambientes computacionais

Como foi possível perceber nas próprias definições anteriores, sistemas de confiança e reputação fundamentam de forma geral relacionamentos entre entidades (seres humanos ou não). Este trabalho focaliza os relacionamentos entre sistemas complexos envolvidos em nuvens computacionais, e nesse contexto mostra-se necessária a existência de mecanismos computacionais para viabilizar as ligações entre as diversas partes envolvidas (*trustee*, *trustor* e elementos intermediários). Para que essas ligações sejam consideradas confiáveis, deve haver, dentro do conjunto de mecanismos, partes responsáveis pela garantia da segurança contra elementos que tenham más intenções. Uma das formas mais utilizadas para evitar problemas causados por esses elementos é a utilização de ferramentas que viabilizem o controle de acesso por meio de processos de autenticação e autorização confiáveis. Essas ferramentas geralmente são fornecidas pelos próprios provedores dos serviços e têm uma única finalidade: protegê-los contra o acesso de usuários não autorizados. Entretanto a proteção dos usuários contra acessos falhos a provedores não confiáveis geralmente permanece como um problema mal resolvido.

Dessa forma cresce a importância de sistemas de confiança e reputação que focalizem não somente a proteção dos provedores de serviços, mas também, e em escala proporcional, dos usuários desses serviços. Essa proteção geralmente é viabilizada, em um primeiro momento, por meio de um acordo legal entre as partes envolvidas que define o escopo de atuação de cada uma, suas obrigações e seus direitos. Após isso são criadas as ligações físicas e lógicas entre os diversos envolvidos considerando os ambientes computacionais e que devem contemplar o fornecimento e utilização de mecanismos para a garantia de segurança às transações. Somente a partir dessa garantia os sistemas computacionais envolvidos poderão ser considerados confiáveis [ABRAMS, 1995].

Essa garantia, entretanto, não é geral. É possível dividi-la em níveis a partir da maior ou menor capacidade de um sistema resistir a ataques de usuários mal intencionados. Dessa forma, apesar de poderem ser utilizados outros parâmetros vinculados à própria reputação das partes envolvidas com a finalidade de qualificar o nível de confiabilidade

(sistema mais ou menos confiável), pode-se dizer que quanto mais seguro um sistema, mais confiável ele será. Nesse contexto, é possível (e até provável) que o nível de confiabilidade de um sistema seja um dos aspectos considerados para a decisão referente à execução ou não de aplicações críticas.

Em relação aos mecanismos de segurança referenciados anteriormente e que são aplicados sobre as ligações físicas e lógicas, estes geralmente utilizam processos criptográficos para viabilizar a garantia do sigilo e da integridade à transmissão das informações, além da própria autenticidade das entidades envolvidas. Cabe salientar que esses serviços de segurança da informação geram tipos de confiança diferentes pois, enquanto a garantia de autenticidade fornece maior ou menor confiabilidade às entidades envolvidas (confiança na identidade), a garantia de sigilo e integridade indica maior ou menor confiabilidade das ligações entre as entidades envolvidas (confiança nas ligações). Existe ainda um terceiro tipo de confiança, o relacionado ao provimento do serviço propriamente dito (confiança no fornecimento do serviço). Este especificamente é derivado dos dois anteriores e da reputação das entidades envolvidas no contexto dos serviços, ou seja, é possível que as entidades confiem mutuamente nas identidades de seus pares, também é possível que as ligações entre as partes sejam consideradas confiáveis e mesmo assim não haja confiabilidade no contexto do fornecimento do serviço tendo em vista aspectos ligados à reputação das partes envolvidas.

Entidades e relações em modelos de confiança

As diversas partes ou entidades envolvidas nos processos que estão baseados em modelos de confiança, como comentado anteriormente, caracterizam-se basicamente por dois papeis: *trustee* (*trusted party*) e *trustor* (*trusting party*). Entretanto é possível que outras entidades intermediárias existam com o objetivo de viabilizar as transações entre os dois papeis principais. Esses papeis podem ser assumidos por diversas entidades reais, dependendo do objetivo do modelo de confiança e principalmente do negócio que está sendo tratado por meio das transações entre as partes.

Em [ETSI TS 133, 2014], por exemplo, é proposto um modelo de confiança que considera a existência de três entidades básicas típicas de um ambiente de MCC: uma rede celular, uma rede sem fio (WLAN) e um usuário. O problema a ser tratado neste exemplo está relacionado à possibilidade de interconexão entre redes sem fio heterogêneas. Nessa

situação específica, a rede celular (elementos de uma rede de comunicações que provê serviços de telefonia móvel celular e considera uma ou mais operadoras, um ou mais domínios e uma ou mais tecnologias de acesso) e a WLAN (entidade composta pelos pontos de acesso e demais elementos de rede que se encontram atrás destes e que fornecem acesso público à rede do padrão IEEE 802.11) fornecem serviços de comunicações de dados ao usuário (elemento que desloca-se entre as redes e é formado pelo próprio assinante das redes celulares e WLAN, pelo equipamento com capacidade de processamento e de comunicação e pelos dispositivos de segurança vinculados – módulos de identificação de usuários, por exemplo).

Nesse caso, é possível considerar que a rede celular e a WLAN sejam as *trusted parties* e o usuário seja a *trusting party*. Note que pelo fato de particularmente as *trusted parties* desse exemplo serem entidades complexas formadas por diversos componentes funcionais distintos (estações rádio base, centros de comutação e controle, centros de autenticação, etc.), também deverão existir modelos de confiança internos, entre os diversos componentes funcionais, para que as redes sejam consideradas confiáveis perante a *trusting party*. Apesar dessas complexidades internas, e provavelmente motivado por elas, não são percebidos elementos intermediários que se interponham entre as *trusted parties* e a *trusting party*.

Outro exemplo de modelo de confiança (Figura 2.4), este versando especificamente sobre transações eletrônicas comerciais, é comentado em [POURSHAHID, 2007]. Nesse caso é sugerida a utilização de uma estrutura que permita a aquisição, por meio da Internet, de livros usados de uma loja virtual de livros (do inglês, *Virtual Book Store* - VBS) especializada nesse tipo de produto. As duas entidades principais, nesse caso, são a VBS e o Comprador (do inglês, *Buyer*). Um modelo de confiança adequado a essa situação deverá considerar, diferentemente do exemplo anterior, alguns elementos intermediários que permitirão a execução das transações, a saber:

- O provedor de acesso à Internet utilizado pelo comprador do livro usado;
- O provedor similar que viabiliza a existência do site a partir do qual são efetuadas as transações eletrônicas;
- O provedor de serviços de pagamento responsável pela transferência de dinheiro da conta bancária (ou cartão de crédito) do comprador para a VBS;

- O provedor de serviços de transporte de mercadorias (no caso, livros usados);
- As diversas redes de comunicações de dados que interligam as entidades anteriormente descritas.

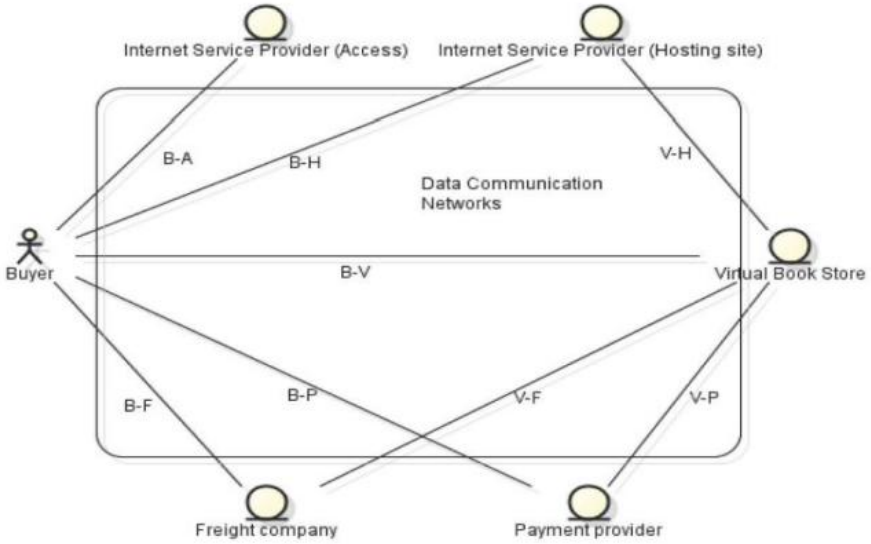

Figura 2.4 - exemplo de modelo de confiança para transações eletrônicas [POURSHAHID, 2007]

Percebe-se que a existência das entidades intermediárias gera um aumento considerável na complexidade do modelo de confiança adequado para a solução do problema. No primeiro exemplo, as relações de confiança restringem-se às ligações entre usuário e rede celular (controlada por um contrato legal de acesso à rede celular), usuário e WLAN (também dependente de um contrato de serviços de acesso à rede de comunicações), e rede celular e WLAN (nesse caso específico, dependendo do nível de acoplamento entre essas entidades, elas podem ter suas ligações controladas por um contrato entre organizações ou, havendo um alto grau de acoplamento, uma entidade poderia pertencer à outra – tipicamente a WLAN poderia pertencer à operadora da rede celular). A quantidade de ligações é pequena e a confiança das entidades em seus pares, como pode ser visto, é basicamente regida por contratos legais.

No segundo exemplo, tendo em vista a quantidade de entidades intermediárias, as relações de confiança tornam-se consideravelmente mais complexas. A Figura 2.4 mostra as diversas entidades e suas ligações. Nesse caso são consideradas apenas as ligações entre a VBS e o *Buyer*, e destes com os elementos intermediários (ligações que envolvem as redes de comunicações de dados, do inglês, *Data Communication Networks*, e ligações

entre elementos intermediários são desprezadas por questões de simplicidade). Note que a relação mais importante desse modelo é aquela que está vinculada diretamente ao negócio tratado (compra e venda de livros usados), ou seja, B-V, a relação entre *Buyer* e VBS. Apesar disso, todas as outras relações interferem em B-V e são importantes para que o negócio seja realizado completamente e satisfaça todas as partes. Assim, a relação B-A deve ser confiável, pois caso não seja, o *Buyer* não poderá confiar no conteúdo que recebe do seu ISP e, consequentemente, poderá não se sentir seguro para efetuar uma transação por meio desse *provider* e que envolva seu dinheiro. Da mesma forma, em B-P e V-P, tanto o *Buyer* quanto a VBS devem confiar no provedor de pagamento (do inglês, *Payment provider*), que é um operador bancário ou de cartão de crédito, e vice-versa, pois é por meio deste que uma entidade paga e a outra recebe o dinheiro. O ISP que é acessado pela VBS e que hospeda seu site também deve ser de confiança para a própria VBS (relação V-H) pois armazena suas informações e viabiliza seus negócios virtuais. Além disso também deve ser digno da confiança do *Buyer* (B-H) pois é desse ISP que virão os dados da VBS com a qual o negócio será realizado. A empresa de frete (do inglês, *Freight company*) também deve ser confiável para o *Buyer* (B-F) e para a VBS (V-F), pois essas entidades esperam que a entrega do produto vendido seja feita de forma satisfatória.

Considerando o ambiente computacional e de comunicações utilizado para as transações do segundo exemplo, pode-se perceber os seguintes tipos de confiança:

- Confiança na identidade – cada entidade deve confiar na identidade do seu par (entidades principais e intermediárias) e isso pode ser feito por meio de mecanismos de autenticação que serão discutidos posteriormente;
- Confiança nas ligações – as entidades não desejam que seus dados sejam modificados (integridade) nem observados por partes não autorizadas (sigilo), o que pode ser viabilizado por meio de técnicas e processos baseados em criptografia;
- Confiança no fornecimento do serviço – cada entidade tem um papel específico no modelo e cada uma é responsável pelo fornecimento de um determinado serviço (venda de livros, transporte de mercadorias, gerenciamento de pagamentos, etc.). Esses serviços devem ser confiáveis para que toda a transação seja concluída com êxito. Como exemplos, considera-se que não seja interessante utilizar uma *Freight company* que tenha fama de danificar os produtos que entrega e não é aconselhável que a VBS hospede seu site em um ISP que não disponha de uma estrutura que

viabilize, de forma satisfatória, a execução de transações eletrônicas (exemplos similares podem ser observados considerando cada outra entidade). Percebe-se que, especificamente para esse tipo de confiança, a reputação das entidades é algo fundamental.

Os dois exemplos comentados anteriormente representam modelos de confiança que são soluções possíveis para problemas específicos. Para outros problemas, os casos devem ser analisados separadamente e demandam a geração de modelos de confiança apropriados.

Confiança em ambientes de computação em nuvem

No contexto da computação em nuvem, os aspectos de confiança são altamente dependentes do modelo de implantação da nuvem computacional, pois a governança sobre os dados e as aplicações dos usuários é delegada para a nuvem (geralmente para o provedor do serviço). Nas arquiteturas tradicionais, a base para a confiança é uma política de segurança eficiente que defina comportamentos permitidos e restrições para o acesso, tanto interno quanto externo, às informações. No caso do modelo de implantação privado (nuvem privada), assegurar a aplicação correta e completa dessa política é algo relativamente simples, considerando que tanto os usuários quanto os provedores do serviço de computação em nuvem fazem parte de uma mesma organização e estão submetidos às políticas gerais desta. Já nos outros modelos de implantação (nuvens pública, comunitária e híbrida), os usuários dos serviços da nuvem tendem a estar fora das organizações responsáveis pelo provimento do serviço, o que torna difícil a imposição, para esses usuários, das políticas de segurança definidas por essas organizações. Por outro lado, ainda nesses últimos modelos, os usuários não têm qualquer controle sobre as políticas dos provedores e necessitam, para poderem utilizar os serviços, submeter-se integralmente aos ditames impostos, mesmo que não concordem totalmente com eles. Este modo de interação gera uma série de riscos para os dois lados, usuários e provedores, pois as partes, apesar de possuírem um documento de acordo (termos e condições ditados pelo provedor), nem sempre estão submetidas à mesma legislação (muitas vezes, o usuário sequer pertence ao mesmo país do provedor).

Apesar de existirem algumas propostas de modelos de confiança [ZISSIS, 2012] [LIN, 2015] [LIN, 2016] que se propõem a minimizar os riscos citados com base em

terceiras partes confiáveis (do inglês, *Trusted Third Party* - TTP) e mecanismos sofisticados de reputação, pode-se considerar que esse problema ainda possui áreas a serem exploradas, principalmente no contexto da MCC.

Confiança e autenticação

Um dos principais objetivos dos mecanismos de autenticação é verificar a identidade de quem solicita a autenticação e posteriormente autorizar a execução de serviços para essa entidade. Nesse contexto, a existência de confiança é um aspecto extremamente relevante para que as transações possam ocorrer, pois as diversas partes envolvidas em uma transação precisam se autenticar umas para as outras a fim de que suas identidades recebam crédito e que haja confiança entre as partes.

Esses aspectos foram discutidos em itens anteriores nos quais foi abordado o conceito de confiança na identidade dos pares que está intimamente ligada aos processos de autenticação utilizados pelos diversos envolvidos. De forma geral, é possível dizer que a autenticação realizada entre duas ou mais entidades que desejem realizar transações eletrônicas cria níveis de confiança na identidade das partes envolvidas [THOMPSON, 2002]. A partir do momento em que um processo de autenticação é desencadeado, é criada uma ligação segura entre os envolvidos (entidades autenticadas) que permite o trânsito confiável (pelo menos no contexto da autenticidade das partes) de informações sem que haja, a princípio, necessidade de novas autenticações. Nesse contexto também devem ser considerados períodos de validade dos processos de autenticação, pois é possível que ocorram problemas durante o intervalo no qual as partes estão trocando informações (ataques de falsificação, por exemplo) que façam com que o processo de autenticação inicial não deva ser considerado permanentemente. Assim, um processo de autenticação executado em um determinado instante garante a confiança na identidade dos pares de uma transação eletrônica por um período limitado, o que gera a necessidade de processos de reautenticação a serem desencadeados posteriormente e que validem as autenticações anteriores. Salienta-se, neste instante, a aplicação da autenticação contínua como forma de minimizar a necessidade de reautenticações, pois a autenticação contínua induzirá a execução de processos constantes de autenticação dos envolvidos que indicarão a confiabilidade das partes durante todos os períodos das transações (sessões).

2.5.2- Pilares da segurança da informação

Integridade

Garantir a integridade de recursos significa assegurar que esses recursos somente serão alterados (modificados ou até deletados) por partes devidamente autorizadas para tal [SUO, 2013]. Nesse contexto, os processos de autenticação, autorização e contabilização (do inglês, *Authentication, Authorization and Accounting* - AAA) são fundamentais para que a integridade das informações seja garantida ou, na pior das hipóteses, seja possível realizar auditorias que mostrem as entidades responsáveis por alterações indevidas nas informações.

Nos ambientes de MCC, pelo fato destes estarem fundamentados em arquiteturas distribuídas, o controle de acesso com base em protocolos e mecanismos de AAA é algo incontestável e deve ser implementado de maneira completa considerando todas a entidades envolvidas nos diversos enlaces necessários para o provimento dos serviços de computação em nuvem.

A Figura 2.2 apresentada anteriormente mostra algumas das entidades integrantes de uma arquitetura de MCC típica e que necessitam ser consideradas nos processos de controle de acesso para que haja garantia da integridade das informações. Uma entidade central nesse contexto é o próprio provedor dos serviços de MCC, pois ele deve ser responsável, perante as demais partes (usuários finais, operadores de rede, provedores de acesso à Internet, dentre outros), pela garantia da segurança das informações, em um contexto amplo, do ambiente de computação em nuvem. Apesar de ter papel central, o provedor dos serviços de MCC também pode estar sujeito a problemas de confiabilidade. Este aspecto será melhor explorado no contexto da confidencialidade das informações constante de itens a seguir.

Disponibilidade

A disponibilidade se refere à propriedade que um recurso deve possuir de estar disponível e em condições de uso por partes autorizadas no momento em que estas necessitem [SUO, 2013].

As arquiteturas de computação em nuvem, de uma forma geral e pelo fato de estarem baseadas em ambientes nos quais há uma grande preocupação com a alta disponibilidade dos recursos (até porque esses recursos devem estar disponíveis para uma gama geralmente grande de usuários com características e necessidades peculiares), favorecem a garantia da disponibilidade dos recursos da nuvem para os usuários.

Confidencialidade

Garantir a confidencialidade das informações significa assegurar que somente partes autorizadas possam acessar essas informações (com qualquer que seja o propósito). Como fica claro pela própria definição, esse serviço de segurança também está fundamentado em protocolos e mecanismos de AAA.

Pelo fato das nuvens computacionais serem ambientes tipicamente de compartilhamento de recursos [RYAN, 2013] por uma gama muito variada de entidades internas (o próprio provedor dos serviços de CC e seus empregados ou outras entidades subcontratadas) e externas à nuvem (os usuários finais, provedores de acesso às redes, etc.), manter a confidencialidade das informações, de uma forma geral, é um dos maiores desafios apresentados nessa área.

No caso específico da MCC, situações claras relacionadas à necessidade de manutenção da confidencialidade das informações podem ser observadas, por exemplo, nos processos de envio das informações dos equipamentos móveis para a nuvem e vice versa, e essas situações devem considerar uma série de possibilidades de autorização para algumas entidades e não autorização para outras.

Para a garantia da confidencialidade, uma possível solução simples para proteger os dados é utilizar criptografia sobre os dados sensíveis aplicando chaves de grupo [ZHOU, 2012a] e enviar os dados para o destino por meio dos mecanismos disponíveis. Somente usuários legítimos (que possuam as chaves de grupo) poderão acessar os dados, entretanto essa solução, considerando o ambiente altamente distribuído sobre o qual a CC baseia-se, demanda uma carga de trabalho excessiva relacionada ao gerenciamento das chaves e pode por em risco, por meio de um único vazamento da chave, o sigilo de todas as informações.

A Confidencialidade Perante o Provedor dos Serviços de CC

Como indicado anteriormente, o problema da garantia da confidencialidade das informações, no ambiente de MCC, é um grande desafio. Esse problema apresenta-se de diversas formas [RYAN, 2013]:

- Garantir a confidencialidade das informações dos usuários móveis em trânsito para a nuvem (e no caminho de volta) perante entidades intermediárias e possíveis atacantes externos;
- Garantir a confidencialidade dessas informações perante outros usuários da própria nuvem computacional;
- Garantir a confidencialidade perante o provedor do serviço de CC.

As duas primeiras situações estão sujeitas a soluções com base em criptografia tradicional (simétrica ou assimétrica) e, além de não serem problemas restritos à CC [RYAN, 2013], podem ser consideradas, nos dias de hoje, de simples solução. A terceira situação, entretanto, apresenta problemas de maior complexidade e será explorada com maior nível de detalhamento a frente neste trabalho.

Como já dito, a criptografia tradicional é uma ótima solução para os problemas de garantia da confidencialidade das informações. Entretanto há situações, como a garantia da confidencialidade das informações dos usuários perante os provedores dos serviços de CC, nas quais não é possível utilizar esse tipo de mecanismo, e essas situações têm sido consideradas um grande desafio de segurança no contexto da computação em nuvem [RYAN, 2013].

As situações indicadas ocorrem, por exemplo, quando aplicações existentes em equipamentos móveis realizam *offloading* de parte de suas execuções para a nuvem. Essas partes que são enviadas necessitam, muitas vezes, de processamento a ser realizado na própria nuvem. A utilização da criptografia convencional para a garantia da confidencialidade, nesse contexto, pode ser feita para o envio das informações e, no máximo, seu armazenamento no ambiente da nuvem. Entretanto o processamento, nessa situação, somente poderá ser realizado caso o conteúdo enviado seja decifrado no próprio ambiente da nuvem. Assim, uma vez no seu formato em claro, os dados a serem

processados estarão sujeitos, também, ao acesso, por vezes indevido, do provedor dos serviços, de seus empregados e até de terceiros subcontratados [RYAN, 2013].

Atualmente, esse problema tem uma resolução parcial por meio de leis, contratos e boas práticas [RYAN, 2013], ou seja, que recai sobre a área das políticas que têm relação com modelos de confiança. Apesar de apresentar resultados favoráveis, essa forma de resolução não é completa considerando que, mesmo com a possibilidade de aplicação de sanções e punições, as políticas podem ser burladas por partes que tenham interesses escusos.

Historicamente esse problema, em contexto mais amplo (sem vinculação específica a um ambiente de nuvem), foi proposto no final da década de 70 e indagava a seguinte questão [HRESTAK, 2014]: "é possível realizar computações sobre dados criptografados sem que, antes, estes sejam decriptografados?"

Muito tempo se passou e somente em 2009 foi publicada uma tese [GENTRY, 2009] na qual foi sugerida uma solução específica para essa questão: um sistema de criptografia totalmente homomórfica. Apesar de o sistema proposto ser mais teórico que prático, o trabalho apresentado em [GENTRY, 2009] inspirou outros trabalhos na área [HELIB, 2014], [LOUK, 2015], [ZKIK, 2016] que têm evoluído no sentido de dar uma solução prática e definitiva para o problema proposto na década de 70 e, por conseguinte, ao problema da garantia da confidencialidade das informações dos usuários das nuvens computacionais perante seus próprios provedores.

Autenticidade, Autenticação e Controle de Acesso

Um dos mais importantes aspectos a serem considerados no contexto da segurança das informações em um ambiente de MCC é a garantia da autenticidade das informações e dos usuários da infraestrutura. Caso um determinado usuário envie para a nuvem, por exemplo, um arquivo a ser compartilhado com outros usuários, é fundamental que haja mecanismos de verificação que permitam a esses outros usuários ter a certeza de que a origem daquela informação realmente é quem aparenta ser.

Dentre as diversas técnicas válidas para viabilizar esse tipo de serviço de segurança destaca-se a assinatura digital criada a partir da utilização de mecanismos de criptografia e Infraestruturas de Chaves Públicas - ICP (do inglês, *Public Key Infrastructures* - PKI).

Vinculada à questão da garantia da autenticidade das informações, e com significado muito próximo a esta, está a possibilidade de autenticação de partes de sistemas entre si. Esse processo consiste na verificação da origem de um acesso, com o objetivo de controlá-lo de forma adequada, determinando se alguém ou alguma coisa realmente é quem ou o que se supõe que seja [WADHWA, 2014]. Nesse sentido, as diversas propostas de processos de autenticação devem focalizar os métodos de autenticação em pelo menos duas situações: o usuário perante a nuvem e a nuvem perante o usuário. Em ambos os casos, esses métodos podem ser baseados na identidade das partes ou no contexto no qual elas se encontram [ALIZADEH, 2016]. Esses assuntos serão explorados em maior nível de detalhamento em itens à frente.

2.5.3- Outros aspectos de segurança envolvendo MCC

Segurança de Aplicações Baseadas em MCC

Garantir a segurança das informações e a privacidade em um ambiente de computação distribuída, como é o caso das nuvens computacionais, é um grande desafio [KOVACHEV, 2011]. No caso específico da privacidade, esta é requerida, no contexto das aplicações voltadas para a CC, para assegurar suas corretas execuções em ambientes isolados e confiáveis, o que inviabilizará a ação de ameaças de uma forma geral.

Assegurar a segurança e a privacidade são aspectos muito importantes para estabelecer e manter a confiança dos usuários móveis em relação ao processamento de aplicações com base em CC [SHIRAZ, 2013]. Nesse contexto, três aspectos se destacam no tocante à segurança das informações utilizadas nas diversas aplicações [ZISSIS, 2012]:

- A segurança do próprio equipamento móvel - prevenção contra invasão no equipamento por meio da ocorrência de artefatos maliciosos (do inglês, *malware*), exploração de vulnerabilidades das aplicações embarcadas ou dos próprios sistemas operacionais e outras ameaças inerentes ao ambiente do próprio usuário antes deste se comunicar com a nuvem propriamente dita;

- A segurança na transmissão das informações do móvel para a nuvem (e vice-versa) - como indicado anteriormente, uma série de entidades se interpõem entre o móvel e a nuvem na qual será provido o serviço de processamento das aplicações, e geralmente essas entidades não necessitam conhecer o conteúdo das informações, mas tão somente retransmiti-las; esse problema torna-se ainda maior a partir do momento em que os meios utilizados para a transmissão das informações são *wireless* (WLAN, redes 3G, 4G, 5G, etc.) e podem ser providos por entidades não tão confiáveis (*Cybercafes*, redes públicas genéricas, etc.), o que sujeita o trânsito das informações a outras vulnerabilidades que dificultam as garantias requeridas, quer no contexto da confidencialidade, da integridade ou até da própria disponibilidade;
- a segurança no processamento e/ou armazenamento interno na nuvem - as CC, por si só, já são ótimos alvos para atacantes diversos pelo fato de concentrarem, em uma única infraestrutura, informações de diversos usuários e de diversas naturezas (inclusive financeiras), aliado a isso e considerando a premissa fundamental do compartilhamento dos ambientes da nuvem que serão utilizados durante o processamento das informações, garantir que as informações que são de propriedade de um determinado usuário sejam protegidas em relação aos outros usuários e até em relação ao próprio provedor dos serviços é um grande desafio e apresenta problemas em aberto.

2.6- CONCLUSÃO

O presente capítulo tem o propósito de apresentar fundamentos teóricos sobre a computação em nuvem, a computação móvel em nuvem e a segurança da informação aplicada nesses contextos.

Os vários temas tratados até aqui indicam algumas das muitas áreas que estão envolvidas no tema principal desta tese: a segurança da informação aplicada em ambientes de MCC. Neste capítulo são apresentados alguns modelos gerais dos ambientes de CC e são tratadas questões importantes no contexto das comunicações que ocorrem em ambientes de CC. Além disso, são explorados conceitos, características e problemas de segurança da informação que envolvem a computação móvel em nuvem.

Como conclusão preliminar, é lícito considerar que as áreas exploradas, além de fundamentais para a vida moderna, apresentam muitos desafios de segurança da informação que merecem ser analisados e resolvidos.

3- FUNDAMENTOS DE AUTENTICAÇÃO E BIOMETRIA

3.1- PROCESSOS CONVENCIONAIS DE AUTENTICAÇÃO

Os processos de autenticação que ocorrem em uma infraestrutura de MCC são ações que viabilizam confiabilidade à identidade das diversas partes envolvidas e permitem a continuidade do provimento dos serviços de uma forma geral. Geralmente o modelo utilizado para permitir que essa autenticação ocorra é estático [CALDERON, 2006]. Esse modelo fundamenta-se em um processo de decisão binário ("autorizado" ou "não autorizado") cujo resultado é válido para um intervalo de tempo determinado (sessão). Esse processo de decisão é baseado, de maneira genérica, em três etapas:

- Inscrição, na qual o usuário suplicante (que desejará autenticar-se posteriormente) firma algum tipo de contrato com a entidade autenticadora (para a qual o suplicante desejará autenticar-se); esse contrato está fundamentado no fornecimento e armazenamento de parâmetros de identificação que serão utilizados posteriormente para validar a autenticação;
- Apresentação, na qual o suplicante solicitará acesso aos recursos disponibilizados pela outra entidade que, por sua vez, solicitará as credenciais do suplicante que serão utilizadas para autenticá-lo;
- Validação, etapa onde é feita a comparação entre os parâmetros apresentados na fase anterior (senha, *token*, característica biométrica, etc.) e aqueles que encontram-se armazenados nas bases de dados adequadas e que foram obtidos e acordados na etapa de inscrição.

O processo de decisão fundamenta-se, então, no resultado da comparação executada na validação. Caso o resultado seja positivo (dados apresentados correspondem aos dados armazenados), o suplicante ganhará acesso aos recursos solicitados, caso seja negativo, o acesso não será concedido.

Para que haja autenticação mútua entre as partes, também será necessário que um processo análogo ao anteriormente descrito seja realizado no sentido oposto (entidade autenticadora para suplicante) ao exemplificado. Essa autenticação mútua, para o contexto da MCC, é algo que cresce muito de complexidade devido à grande quantidade de entidades envolvidas. Observando novamente a Figura 2.2, percebemos a necessidade de

autenticação, por exemplo, entre equipamentos móveis (do inglês, *Mobile Devices*), pontos de acesso (do inglês, *Access Points*), Estações Radio Base - ERB, servidores, controladores da nuvem (do inglês, *cloud controllers*), *data centers*, etc.

Cabe salientar que as credenciais citadas anteriormente podem ser de três categorias [SYTA, 2010] :

- Algo que o usuário conhece (baseado no conhecimento, do inglês, *Knowledge-based*);
- Algo que o usuário possui (baseado na posse, do inglês *Possesion-based*);
- Algo que é inerente ou intrínseco ao usuário (baseado em biometria, do inglês, *Biometric-based*).

A Figura 3.1 ilustra as categorias indicadas e exemplifica cada uma delas com alguns métodos citados anteriormente e com outros que serão explorados à frente.

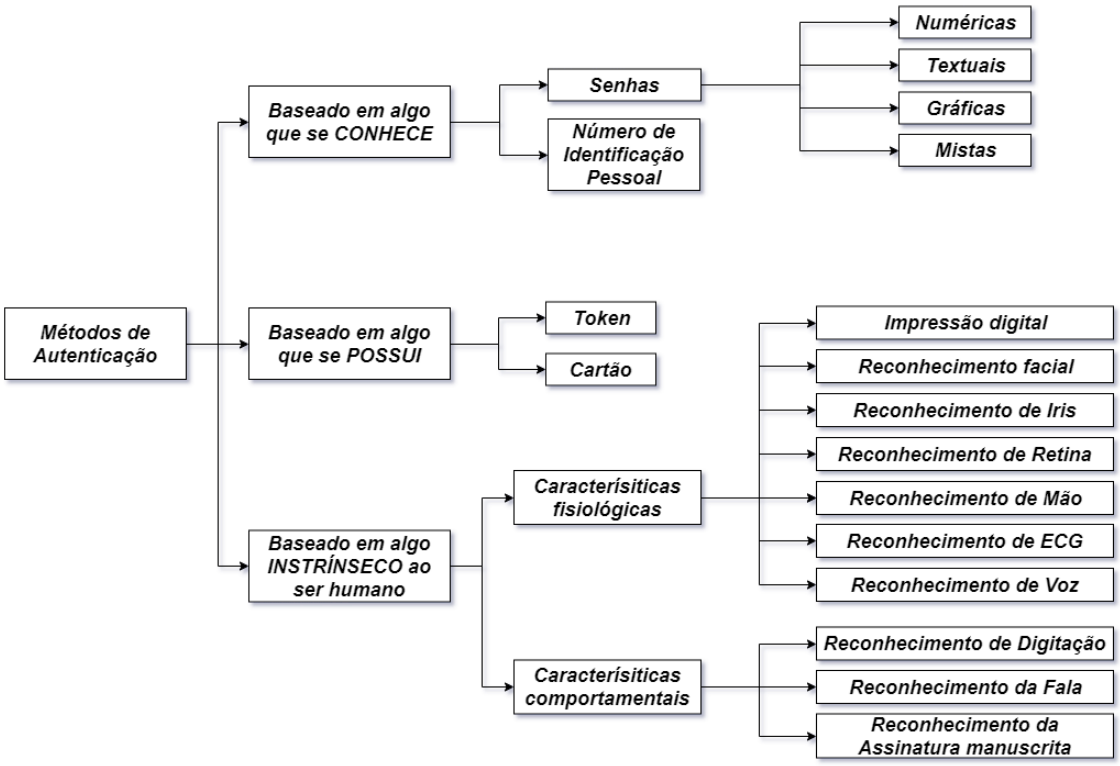

Figura 3.1 - Classificação de métodos de autenticação

São explorados, a seguir, alguns esquemas típicos de autenticação.

3.1.1- Usuário / Senha

Uma dos processos de autenticação mais utilizados é aquele baseado no emprego dos parâmetros "usuário" - dado ostensivo - e "senha" (do inglês, *password*) - dado secreto - conhecida apenas pelas partes autorizadas, que são credenciais da primeira categoria (algo que o usuário conhece). Esses parâmetros devem ser criados na fase de inscrição e utilizados na fase de apresentação para que o usuário autentique-se à entidade (e vice-versa, quando for o caso).

Apesar de largamente utilizado, o mecanismo de autenticação baseado em usuário e senha possui diversas fragilidades [FURNELL, 2000], [SAUVER, 2009] como, por exemplo:

- Problemas na criação das senhas – em situações nas quais é concedido ao usuário a possibilidade de escolher a própria senha, é comum que sejam escolhidos termos (numéricos ou alfanuméricos) que tenham algum significado particular para o usuário (datas de aniversários, nomes de pessoas conhecidas, etc.) ou que sejam palavras existentes na língua do indivíduo; em ambos os casos, essas senhas estão sujeitas a ataques de busca por exaustão baseados em dicionários conhecidos (vinculados ou não ao indivíduo) [SCHNEIDER, 1996], o que diminui consideravelmente a segurança da senha escolhida; uma forma de minimizar esse problema é impor ao indivíduo a utilização de uma mistura de letras, números e caracteres especiais para a criação de senhas que não façam parte de quaisquer dicionários (existentes ou criados pelo atacante a partir do conhecimento sobre o indivíduo); outra solução para esse problema baseia-se no fornecimento, para o usuário, de uma senha criada aleatoriamente (ou pseudo aleatoriamente) pelo próprio sistema responsável pelo cadastro dos parâmetros de identificação do usuário; nas duas soluções indicadas surge o inconveniente da difícil memorização da senha, o que é solucionado com a anotação da senha em algum local protegido, o que também é um problema pois esse local pode não ser tão protegido e a senha pode ser encontrada por um usuário mal intencionado, o que certamente gerará uma autenticação falsa;

- Problemas no armazenamento dos parâmetros do usuário – isso é observado quando especificamente a senha do usuário é armazenada em claro na base de dados da entidade autenticadora; nesse caso, um intruso que consiga acesso a essa base de dados poderá capturar os dados de autenticação do usuário e passará a ter capacidade de autenticar-se como se fosse o próprio usuário; uma forma de minimizar esse tipo de problema é armazenar parâmetros relacionados à senha (tipicamente o *hash*) ao invés da própria senha; assim, caso o intruso tenha acesso à base de dados, ele não obterá a senha em claro do usuário;
- Problemas na transmissão das informações entre usuário e entidade autenticadora – para que o usuário realize sua autenticação, ele precisará enviar suas informações para o autenticador e, caso o canal de comunicações existente entre o usuário e a entidade não esteja protegido (física ou logicamente), haverá a possibilidade de um intruso capturar os dados do usuário (tipicamente o parâmetro sensível - a senha) e utilizá-los posteriormente para realizar uma autenticação falsa; uma solução para esse problema é enviar, no lugar da senha, parâmetros dinâmicos (sincronizados com o autenticador) calculados a partir da senha e que não permitam, quando capturados por um intruso, a dedução desta; nesse caso deve ser possível realizar o cálculo desses parâmetros também no ambiente do autenticador para que ele possa validar a autenticação do usuário.

3.1.2- Número de Identificação Pessoal

Outra forma de autenticação, semelhante à baseada em usuário e senha, está fundamentada na utilização de um Número de Identificação Pessoal (do inglês, *Personal Identification Number* - PIN) vinculado a parâmetros secretos armazenados pelo usuário (geralmente em um dispositivo de armazenamento removível) e pela entidade autenticadora. Esse método mistura credenciais da primeira (algo que o usuário conhece) e da segunda (algo que o usuário possui) categorias e é tipicamente utilizado para a autenticação de equipamentos perante servidores de acesso a redes de comunicações (autenticação de *smartphones* em suas redes de comunicações, por exemplo) e, apesar de apresentar algumas diferenças em relação ao método baseado em dados de usuário e senha, é equivalente a este.

O PIN, particularmente no contexto dos telefones celulares, pode ser aplicado para a autenticação tanto do aparelho quanto do próprio usuário da rede celular (Módulo de Identidade do Assinante, do inglês, *Subscriber Identity Module* - SIM) [FURNELL, 2008] e, apesar de ser o mecanismo mais comum para a autenticação de equipamentos móveis, não pode ser considerado muito seguro para a autenticação de usuários [BABAEIZADEH, 2014].

Nesse método, o usuário compartilha parâmetros secretos com a rede de comunicação móvel celular, e esses parâmetros encontram-se em um dispositivo removível (SIM-card) que fica com o usuário (geralmente inserido em seu equipamento) e em um elemento da rede celular (tipicamente o centro de autenticação). O processo de autenticação pode ocorrer de forma semelhante ao tratado anteriormente, e por esse motivo, muitos dos problemas observados no processo baseado em usuário e senha também ocorrem na autenticação que tem como base a utilização do PIN.

3.1.3- Token

Um *token* de autenticação é um equipamento que armazena informações que podem ser utilizadas por um usuário para provar sua identidade perante uma entidade autenticadora e é geralmente pequeno o suficiente para ser carregado sem dificuldades (no bolso do indivíduo, por exemplo). Ele faz parte da segunda categoria de credenciais (algo que o usuário possui) e é utilizado, grande parte das vezes, para a criação de senhas ou chaves dinâmicas de forma sincronizada entre o próprio token e um mecanismo (hardware ou software) vinculado à entidade autenticadora [GORMAN, 2003]. A cada momento em que o usuário necessita de uma senha para realizar uma autenticação, o *token* cria esse parâmetro levando em consideração dados e algoritmos que também são de conhecimento da entidade autenticadora, dessa forma a mesma senha será criada nos dois lados da autenticação e permitirá a validação da identidade do detentor do *token*. O principal problema desse processo está ligado à perda ou extravio do *token*. Nestas situações, o indivíduo que encontrar ou obtiver esse equipamento poderá autenticar-se perante a entidade autenticadora como o usuário autêntico (dono do *token*). Por esse motivo, geralmente a senha dinâmica criada pelo *token* é utilizada em conjunto com senhas estáticas do usuário autêntico, o que confere maior confiabilidade e segurança ao processo como um todo a partir do momento em que devem ser utilizadas duas categorias de

credencias para a autenticação: algo que o usuário possui (o *token*) e algo que o usuário conhece (a senha estática).

Tokens também podem ser utilizados para o armazenamento de certificados digitais. Dessa forma eles passam a ser semelhantes a *smart cards* que armazenam certificados digitais e realizam processamentos criptográficos (simétricos e assimétricos) que conferem segurança às transações.

3.1.4- Certificado Digital

Um certificado digital é uma estrutura de dados que contém valores de chave pública (criptografia assimétrica) e um conjunto de informações de identificação da entidade a qual essa chave está vinculada [RFC 5280, 2008]. Essa vinculação é feita e assinada tipicamente por autoridades certificadoras que fazem parte de uma Infraestrutura de Chaves Públicas (do inglês, *Public Key Infraestruture* - PKI). A confiabilidade do certificado deriva da assinatura feita pela autoridade certificadora e que está contida no certificado. Nesse contexto observa-se o princípio da transitividade da confiança, pois o certificado só é considerado confiável a partir do momento em que a autoridade certificadora que o assinou é uma entidade confiável ("confio na autoridade, logo confio no certificado assinado por ela"). Essa ideia pode ser estendida para outras autoridades certificadoras integrantes da PKI e que assinam os certificados de autoridades de níveis mais baixos (que assinam certificados de entidades que não são autoridades, por exemplo), e assim por diante até atingir-se a autoridade certificadora raiz da PKI. Dessa forma considera-se que caso haja uma autoridade certificadora confiável na cadeia de certificação de um certificado, este certificado será considerado confiável.

Outro modelo que pode ser utilizado para conferir confiabilidade ao certificado digital é o da *web of trust*. Neste caso o certificado é autoassinado (assinado pelo próprio detentor) ou assinado por usuários que atestam a validade do certificado que estão assinando. Em ambos os casos (PKI e *web of trust*), a confiabilidade do certificado é viabilizada por terceiras partes confiáveis (do inglês, *Trusted Third Party* - TTP) que assinam o certificado validando sua autenticidade.

Certificados digitais podem ser utilizados de várias formas, para viabilizar a autenticação de usuários (pessoas) e equipamentos (servidores web, servidores de email,

controladores de nuvem - do inglês, *cloud controllers*, centros de dados - do inglês, *data centers*, etc.). Também é comum utilizar-se, por exemplo, em assinaturas digitais para atribuir autenticidade a documentos e em processos de *login* em substituição aos parâmetros usuário e senha. No contexto do acesso às MCC, em grande parte das vezes, os certificados digitais têm o papel fundamental de fornecer autenticidade às partes envolvidas (ou pelo menos a uma das partes) e viabilizar chaves criptográficas que permitem a garantia do sigilo e da integridade das informações trocadas entre os envolvidos (protocolos de camada de soquete seguro, do inglês, *Secure Socket Layer* - SSL, de segurança da camada de transporte, do inglês, *Transport Layer Security* - TLS [RFC 5246, 2008] e protocolo de transferência de hipertexto seguro, do inglês, *Hyper Text Transfer Protocol Secure* - HTTPS [RFC 2818, 2000], por exemplo).

Pelo fato de serem estruturas de dados lógicas, os certificados podem ser armazenados em diversos locais: na própria memória do computador (*hard drive*), em mídias removíveis (*pen drive*, etc.), em *tokens* e *smart cards*. O certificado possui somente dados ostensivos, dessa forma pode ser transmitido livremente pois sua integridade e autenticidade permanecem garantidas por meio da assinatura da entidade que o criou. Vinculado ao certificado digital, o detentor também possui uma chave privada (relacionada à chave pública contida no certificado) que geralmente fica protegida física e logicamente (ou pelo menos logicamente) contra acessos não autorizados. Dessa forma, para que uma chave privada seja utilizada, é necessário que o usuário a possua (2ª categoria de credenciais – algo que se possui) e que saiba a senha (geralmente uma frase senha) utilizada para sua proteção lógica (1ª categoria de credenciais – algo que se conhece).

No caso específico da MCC, os certificados podem ser utilizados, de maneira bastante eficaz, na autenticação dos provedores de serviços de CC perante os seus usuários (autenticação fim a fim da parte que fornece o serviço).

3.2- ASPECTOS DE BIOMETRIA

Antes de explorar os aspectos conceituais da biometria, merece menção o fato dela ter, como foco no contexto deste trabalho, a autenticação de usuários perante os sistemas que estes desejam utilizar. Para a MCC, essa técnica pode ser utilizada para viabilizar a autenticação do ser humano que se encontra atrás da interface (ME) que acessará os serviços desejados da CC, o que caracteriza um tipo de autenticação fim a fim.

3.2.1- Aspectos Gerais

O termo biometria pode ser compreendido, em um sentido amplo, como o estudo ou a medição de características inerentes aos seres vivos, entretanto essa definição é muito genérica. De maneira precisa, a biometria pode ser compreendida como [RIERA, 2009] uma ciência baseada em mecanismos automáticos de análise de aspectos humanos que tem como objetivo realizar o reconhecimento de um indivíduo ou diferenciá-lo de seus semelhantes com base em características fisiológicas e comportamentais.

Vinculada à definição de biometria, surge também o conceito de característica biométrica, que pode ser entendida como um aspecto mensurável e único de um indivíduo que pode ser utilizado para seu reconhecimento ou para a verificação de sua identidade, ambos de forma automática.

Esse último conceito considera alguns requisitos chave que merecem maior detalhamento. O primeiro deles diz respeito ao fato da biometria explorar características únicas. Esse aspecto indica que uma dada característica, quando explorada em mais de um indivíduo, deve apresentar diferenças em níveis adequados ao contexto da análise considerada. Dessa forma, essa característica poderá ser utilizada como meio de diferenciação entre elementos semelhantes.

Outro requisito vinculado ao conceito explorado indica que as características biométricas devem ser mensuráveis, ou seja, devem existir formas práticas de medição dessas características de tal forma que elas possam ser quantificadas e representadas por meio de parâmetros objetivos.

As capacidades de reconhecimento e de verificação de identidade automáticas também devem ser aspectos viabilizados por meio da biometria. No primeiro caso, as medições realizadas sobre as características únicas devem ser suficientes para que o indivíduo seja reconhecido sem a necessidade de outras análises. No segundo contexto, essas mesmas medições devem permitir que um indivíduo tenha a sua identidade verificada a partir da comparação entre os parâmetros medidos e os armazenados em bases de identificação.

Cabe salientar que essas características únicas que são apresentadas pelos indivíduos e que podem ser medidas e quantificadas servido para diferenciar indivíduos semelhantes são exatamente o que define uma das categorias de credenciais utilizadas em processos de autenticação (algo que é inerente ao usuário e faz parte da sua identidade).

Nesse contexto, percebe-se que os métodos de biometria podem ser utilizados tanto como processos de identificação (reconhecimento de quem supostamente o indivíduo é) quanto de autenticação (validação da suposta identidade).

Para que a biometria gere resultados objetivos, geralmente são realizadas comparações entre os parâmetros medidos no instante da autenticação e aqueles previamente armazenados em bases de dados criadas a partir do cadastramento das características dos indivíduos [CLARKE, 2002a]. Essas comparações nem sempre são perfeitas (na verdade, quase nunca há semelhança total entre a característica medida e a armazenada) e para que a resposta do sistema biométrico possa ser binária (autêntico ou não autêntico), é necessário que sejam escolhidos níveis de precisão considerados suficientes para a autenticação adequada do indivíduo. Dessa forma, caso seja escolhido um nível alto de precisão (maior proximidade entre o que é medido e o que está armazenado) e o indivíduo seja reconhecido, a certeza de que o processo gerou um resultado correto será grande. Entretanto o aumento desse nível também gera um problema: a maior probabilidade de que o indivíduo não seja reconhecido. É necessário, então que sejam escolhidos níveis intermediários (nem altos nem baixos) coerentes com o processo de reconhecimento desejado. Geralmente esses níveis são escolhidos de forma proporcional aos níveis de criticidade dos recursos cujos controles de acesso estão sendo viabilizados por meio dos processos biométricos.

Métricas de análise aplicadas à biometria

Vinculados a essas ideias, são empregados os conceitos de taxa de aceitação verdadeira (do inglês, *True Acceptance Rate* - TAR), taxa de aceitação falsa (do inglês, *False Acceptance Rate* - FAR), taxa de rejeição falsa (do inglês, *False Rejection Rate* - FRR) e taxa de igualdade de erros (do inglês, *Equal Error Rate* - EER) [ANAND, 2010], [CLARKE, 2002a], que são métricas de análise comumente aplicadas aos processos que envolvem biometria. A TAR e a FAR representam, respectivamente, a taxa de aceitações corretas e a taxa de aceitações equivocadas do sistema. A primeira indica a quantidade

relativa de indivíduos reconhecidos corretamente pelo sistema pelo fato dos dados analisados corresponderem a dados presentes na base de identificação, e a segunda a quantidade relativa de indivíduos reconhecidos pelo sistema sem que seus dados estejam de fato nas bases de dados de identificação. A FRR sinaliza uma situação oposta, isto é, a taxa de rejeição (não reconhecimento) de indivíduos cujos dados encontram-se nas bases comentadas. A EER representa o valor percentual vinculado à taxa na qual a FAR e a FRR terão mesmo valor.

Essas três últimas grandezas estão representadas graficamente na Figura 3.2. Nela percebe-se que, em situações nas quais o nível de acurácia para o processo de reconhecimento é pequeno (pequenos valores de x), são observadas altas taxas de aceitação errada (FAR) e baixas taxas de rejeição equivocada (FRR). De forma contrária, caso o nível de acurácia escolhido seja grande, serão percebidas altas FRR e baixas FAR.

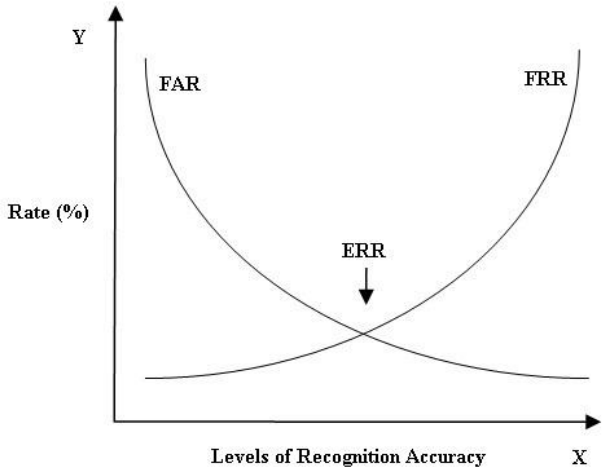

Figura 3.2 – FAR X FRR
(adaptado de [CLARKE, 2002a])

Além das taxas indicadas anteriormente, o presente trabalho utiliza algumas outras métricas [METZ, 1978] vinculadas à estatística e cujas definições se baseiam nas quantidades de:

- **verdadeiros positivos** (do inglês, *true positives* - *tp*) - classificação como positivo de algo que realmente é positivo;
- **verdadeiros negativos** (do inglês, *true negatives* - *tn*) - classificação como

negativo de algo que realmente é negativo;

- **falsos positivos** (do inglês, *false positives* - *fp*) - classificação como positivo de algo que deveria ser negativo;
- **falsos negativos** (do inglês, *false negatives* - *fn*) - classificação como negativo de algo que deveria ser positivo;

Essas métricas são a **acurácia** (do inglês, *accuracy* - *Acc*), a **precisão** (do inglês, *precision* - *Pre*), a **sensibilidade** (do inglês, *sensitivity* - *Sen*), e o **escore F1** (do inglês, *F1-score* - *F1s*), representadas pelas equações (3.1), (3.2), (3.3) e (3.4) , respectivamente.

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{3.1}$$

$$Pre = \frac{tp}{tp + fp} \tag{3.2}$$

$$Sen = \frac{tp}{tp + fn} \tag{3.3}$$

$$F1s = 2 \cdot \frac{Pre \cdot Sen}{Pre + Sen} \tag{3.4}$$

De maneira geral, as métricas indicadas têm os seguintes significados:

- Acurácia: indica o desempenho geral de um modelo e representa, dentre todas as classificações, quantas o modelo classificou corretamente;
- Precisão: em um dado modelo, dentre todas as classificações positivas (verdadeiras ou falsas), quantas estão corretas (ou são verdadeiras);
- Sensibilidade: em um dado modelo, dentre todas as classificações que deveriam ser positivas, quantas de fato são;
- *F1-Score*: média harmônica entre a precisão e a sensibilidade.

3.2.2- Tipos de biometria

De uma forma geral, a biometria pode ser dividida em duas categorias ou tipos [CLARKE, 2002b], [RAHAL, 2007]: a biometria fisiológica e a comportamental. Essas categorias serão exploradas de forma detalhada nos próximos itens.

Biometria fisiológica

Essa categoria de biometria está baseada na medição de aspectos fisiológicos percebidos no corpo humano. Salienta-se que, como indicado, para que esses aspectos sejam considerados características biométricas, eles precisarão atender aos requisitos enunciados anteriormente.

Alguns exemplos de biometria fisiológica serão descritos a seguir.

Reconhecimento da impressão digital (do inglês, *fingerprint recognition*)

Essa é uma das formas de biometria mais utilizadas. Ela viabiliza o controle de acesso a inúmeros recursos (instalações, computadores, telefones, etc.) e encontra-se em um estágio bastante consolidado de seu desenvolvimento [CLARKE, 2002a], [JAIN, 2004], [RAHAL, 2007], [ANAND, 2010].

Esse tipo de reconhecimento baseia-se na verificação dos padrões vinculados ao posicionamento das linhas criadas pelas alternâncias dos *ridges* e *valleys* das digitais dos indivíduos (Figura 3.3). Esses padrões apresentam diferenças aparentes quando são feitas comparações entre indivíduos diferentes.

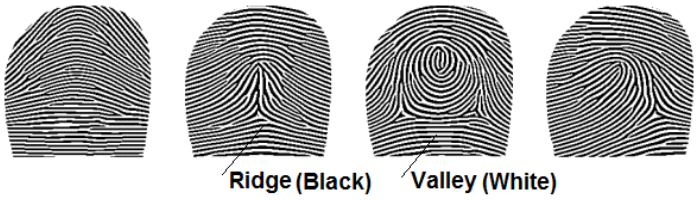

Figura 3.3 - Impressões digitais

Reconhecimento facial (do inglês, *facial recognition*)

Esse método de reconhecimento biométrico baseia-se nas diferentes medidas e características que podem ser encontradas em um rosto humano [CLARKE, 2002a], [JAIN,

2004], [RAHAL, 2007], [ANAND, 2010]. Geralmente os parâmetros escolhidos para representar uma face são aqueles que não mudam (ou mudam pouco) com o passar do tempo. São exemplos desses parâmetros, medidas retiradas de áreas próximas às bochechas, aos olhos, ao nariz e à boca que permitem identificar seus formatos. O processo de reconhecimento é subdividido nas fases de captura da imagem (feita geralmente por meio de uma câmera fotográfica ou de vídeo), processamento (medição dos parâmetros escolhidos previamente), comparação (dados medidos X dados armazenados) e tomada de decisão (havendo coincidências suficientes entre os dados comparados, o indivíduo é considerado autêntico).

Uma vantagem relevante desse método em relação a outros mecanismos de biometria fisiológica é o fato de ele poder ser não intrusivo. Isso é possível porque todas as suas fases podem ser executadas sem a intervenção explícita no usuário.

A Figura 3.4 mostra algumas possíveis medidas que podem ser capturadas e utilizadas no processo de reconhecimento facial.

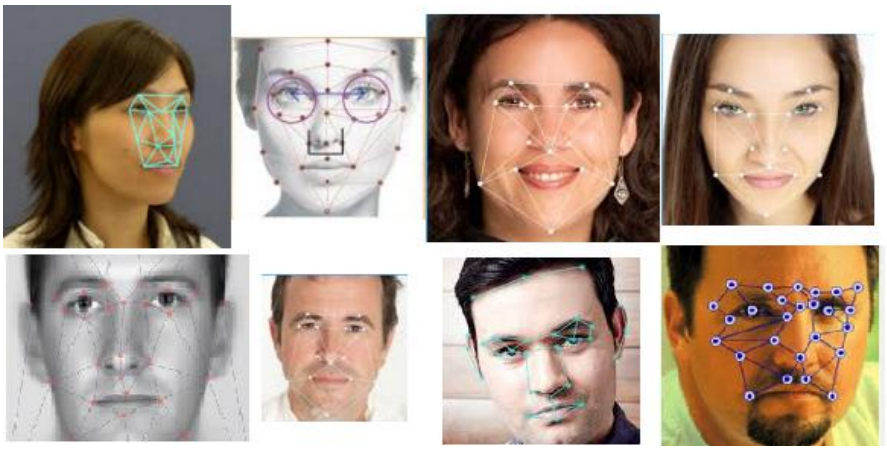

Figura 3.4 – medição de parâmetros da face

Reconhecimento da forma da mão (do inglês, *handshape recognition*)

Esse método de biometria fisiológica baseia-se na medição de distâncias e formatos observados nas mãos das pessoas [JAIN, 2004], [RAHAL, 2007]. Geralmente essas medições são feitas a partir da visualização da palma da mão e focalizam características como comprimento e largura dos dedos e as distâncias entre suas juntas.

Semelhante ao processo anterior, o reconhecimento por meio da geometria da mão também é feito a partir da captura dos parâmetros desejados realizada por equipamentos adequados. A Figura 3.5 mostra alguns exemplos de equipamentos utilizados para a captura das imagens e de resultados do processo de digitalização do formato da mão.

Note que os equipamentos constantes da figura, além de viabilizarem a captura dos parâmetros vinculados especificamente à mão do indivíduo, permitem também a entrada de dados por meio da digitação em um teclado. Dessa forma esses equipamentos podem ser utilizados para a realização de processos de autenticação multimodal baseada em biometria (no caso o reconhecimento a partir do formato da mão) e na inserção de parâmetros secretos (senhas, por exemplo).

Figura 3.5 – Procedimento de reconhecimento da geometria da mão

Reconhecimento do olho, da iris ou da retina) (do inglês, *eye*, *iris or retina recognition*)

A primeira opção baseia-se na Iris. Essa forma de reconhecimento biométrico viabiliza a identificação e autenticação de uma pessoa a partir de características específicas encontradas no olho humano. Acredita-se que essas características permaneçam imutáveis durante toda a vida do indivíduo, o que é um aspecto bastante interessante em se tratando de características fisiológicas do ser humano [CLARKE, 2002a], [JAIN, 2004], [RAHAL, 2007], [ANAND, 2010].

Um detalhe que faz com que esse processo seja considerado um dos mais precisos de sua categoria é o fato da probabilidade de existirem duas iris iguais ser extremamente pequena (da ordem de 10^{-78}).

Dentre algumas características específicas que são alvo dos processos de medição aplicados sobre a iris, considera-se que o principal parâmetro a ser analisado é a *trabecular meshwork*, que é uma malha complexa semelhante a um teia de aranha e formada por tecidos enrugados.

Tendo em vista a precisão bastante grande dessa característica fisiológica, é necessário que a captura da imagem da Iris seja feita por equipamentos que permitam altas resoluções.

A segunda opção está baseada na Retina, que está localizada na parte anterior do olho humano e, diferentemente dos elementos fisiológicos analisados anteriormente, não é visível a olho nu. Essa área do olho humano possui uma grande quantidade de vasos sanguíneos que formam uma teia complexa que pode ser percebida caso a parte de trás de um olho seja observada. Tendo em vista o fato de ser uma área de difícil registro (captura de sua imagem), devem ser utilizados equipamentos especiais e é necessário que haja completa imobilidade do olho analisado para que não sejam gerados erros na captura. Semelhante ao processo baseado na Iris, o reconhecimento da retina também é considerado um dos mais precisos de sua categoria.

A Figura 3.6 mostra a Iris de um olho humano e a sua *trabecular meshwork*, além de também apresentar retinas com suas teias de vasos sanguíneos que formam um padrão bastante específico.

Figura 3.6 – Iris e Retina

Reconhecimento de sinais cardíacos

A biometria baseada em sinais cardíacos é única (possível de ser distinguida entre pessoas diferentes), mensurável, independe da vontade dos indivíduos, segura (difícil de

ser falsificada) e está presente em todos os indivíduos que estão vivos [LIN, 2017]. Dessa forma mostra-se como uma boa opção para ser utilizada como método de autenticação de indivíduos em sistemas complexos e, por esses, dentre outros motivos, é o tipo de biometria escolhida para o protocolo proposto nesta tese.

Considerando a relevância desse tema para o presente trabalho, esse item será explorado com um nível de detalhamento maior no Capítulo 4-Fundamentos de Autenticação baseada em Eletrocardiografia e de Aprendizado de máquina.

Biometria comportamental

Essa categoria está fundamentada na medição de aspectos comportamentais percebidos nas ações dos seres humanos. Da mesma forma que na biometria fisiológica, para que esses aspectos sejam considerados características biométricas, eles precisarão atender aos requisitos enunciados anteriormente (características únicas e mensuráveis, capacidades automáticas de reconhecimento e verificação de identidade).

Alguns exemplos de biometria comportamental serão descritos a seguir.

Análise da dinâmica de digitação (do inglês, *Keystroke Analyses*)

Essa técnica de reconhecimento biométrico é baseada na forma utilizada pelas pessoas para realizar digitações utilizando teclados [CLARKE, 2002a], [CLARKE, 2007], [JAIN, 2004], [RAHAL, 2007]. Esse método não é tão preciso quanto os processos vinculados a características fisiológicas e é pouco provável que um padrão de digitação seja considerado único, entretanto essa técnica oferece uma quantidade de informações suficientes para permitir a identificação provável de um indivíduo.

A medição típica realizada nesse caso focaliza os diferentes intervalos observados entre *keystrokes* sucessivas. Essas medições podem ser realizadas de forma intrusiva em momentos imediatamente anteriores à autenticação do indivíduo (tipicamente o usuário deve digitar um texto predefinido cujo padrão encontra-se armazenado nas bases de dados de identificação) ou de maneira não intrusiva durante um processo de digitação normal do indivíduo (nesse caso não há textos predefinidos e as medições são feitas de forma transparente para o usuário). Apesar da análise comparativa entre padrões medidos e

armazenados ser mais simples no primeiro caso, a segunda forma geralmente não é descartada e as duas maneiras são executadas sequencialmente.

Um aspecto bastante positivo desse método que favorece a sua utilização para prover autenticação não intrusiva e transparente para os usuários é a não necessidade de serem utilizados equipamentos específicos para a medição das características biométricas (utiliza-se unicamente o teclado no qual o indivíduo está habituado a digitar). Dessa forma, apesar da precisão reduzida em relação a outros métodos biométricos, a possibilidade de captura e análise de amostras extensas de *keystrokes* pode viabilizar, sem que o usuário sequer perceba, taxas adequadas de autenticação correta.

Análise da dinâmica de movimento do Mouse

Esse método está baseado na detecção dos padrões individuais de movimentação do mouse para a realização de tarefas específicas [PUSARA, 2004], [GOUDELIS, 2005]. Algumas implementações desse tipo de reconhecimento biométrico utilizam técnicas de inteligência artificial (AI) para aprender o padrão particular de um indivíduo e para compará-lo com a movimentação capturada em um determinado momento de análise. Essas técnicas de AI realizam seus aprendizados a partir, por exemplo, de medições de velocidade de movimentações e de frequência de cliques de botões.

Esse processo não é muito preciso e geralmente é utilizado em conjunto com outros métodos similares (*keystroke analyses*, por exemplo) para que a taxas de erros (falsas aceitações e falsas rejeições) sejam diminuídas. Da mesma forma que o anterior, esse método também pode ser utilizado facilmente para fornecer autenticação não intrusiva e transparente para os usuários.

Reconhecimento de Voz

Esse método está fundamentado em aspectos da voz de um indivíduo que abrangem tanto características físicas (timbre de voz) quanto comportamentais (modo de falar) [CLARKE, 2002a], [JAIN, 2004], [RAHAL, 2007]. Dessa forma, em algumas situações esse processo pode ser considerado como uma técnica de biometria fisiológica, em outros momentos pode ser visualizado como um método de biometria comportamental e, em outros instantes, como ambos.

Esta modalidade biométrica é uma das mais difundidas sendo utilizada não apenas como modo de autenticação, mas também (e talvez principalmente) em aplicações de reconhecimento de voz para execução de comandos remotos e para a conversão da fala em texto escrito.

Para situações de autenticação, padrões das ondas sonoras emitidas (frequência, amplitude, etc.) pelo indivíduo são capturados por equipamentos específicos (microfones) e comparados aos elementos armazenados nas bases de dados de identificação. Cabe salientar que essas bases de identificação são criadas a partir da captura prévia de amostras relevantes dos padrões dos sons emitidos pelos usuários que deverão ser autenticados no futuro.

Reconhecimento de assinatura manuscrita

Esse processo de biometria visa ao reconhecimento da assinatura manuscrita criada por um indivíduo [CLARKE, 2002a], [JAIN, 2004], [RAHAL, 2007], [ANAND, 2010]. Para que esse padrão de escrita seja capturado, podem ser utilizados recursos como telas sensíveis ao toque, scanners e tablados especiais criados especificamente para esse fim. O reconhecimento de assinatura baseia-se em aspectos mensuráveis obtidos geralmente a partir da análise da imagem da assinatura. Nesse contexto, certas vezes a forma de aquisição da imagem não é relevante, mas somente a própria assinatura digitalizada. Entretanto, existem implementações mais sofisticadas desse método que também consideram outros aspectos (tempo que o indivíduo leva para assinar, pressão exercida pelo instrumento de assinatura e inclinação do mesmo em relação à superfície de captura, por exemplo) e que demandam a utilização obrigatória de mecanismos de captura específicos que fornecem, além da própria imagem da assinatura, os outros parâmetros considerados.

Como nos métodos anteriores, o padrões adquiridos também deverão ser comparados aos existentes nas bases de identificação para que o indivíduo seja autenticado.

3.3- AUTENTICAÇÃO BASEADA EM BIOMETRIA

Como indicado anteriormente, a principal característica que diferencia os mecanismos de biometria dos demais processos de autenticação é a dependência de aspectos intrínsecos ao ser humano. Assim, apesar das vantagens, fica óbvio que a biometria aplicada aos ambientes de MCC viabiliza a autenticação de apenas uma das várias partes integrantes da infraestrutura: o usuário.

Já tendo abordado, anteriormente neste trabalho, diversas formas de biometria, a seguir serão focalizados alguns tipos dessa técnica que mostram-se bastante adequados aos processos de autenticação de usuários em ambientes de MCC

3.3.1- Autenticação baseada em impressão digital

Provavelmente a forma mais largamente utilizada no mundo para a autenticação biométrica genérica de usuários, o processo baseado no reconhecimento da impressão digital dos indivíduos pode ser considerado o mais usual também no contexto dos ambientes de MCC.

Além de existirem vários mecanismos de leitura de impressão digital atualmente aceitos comercialmente, muitos desses mecanismos têm sido incorporados aos próprios aparelhos móveis e é muito comum, hoje em dia, ao adquirir notebooks ou telefones inteligentes (do inglês, *smartphones*), escolher modelos que possuam esses leitores embutidos, pois há uma grande quantidade de equipamentos desse tipo disponíveis no mercado. A Figura 3.7 exemplifica modelos de equipamentos móveis que possuem leitores de *fingerprints*.

Figura 3.7 – Dispositivos móveis com leitores de impressão digital

Nesse sentido, e coerente com a realidade do mercado, existem várias propostas de mecanismos de autenticação baseados em *fingerprint* para controle de acesso aos serviços de MCC [HAMAMI, 2015], [RAJA, 2013], [WANG, 2011] e todas elas, apesar de possuírem peculiaridades específicas, indicam as necessidades citadas anteriormente de inscrição, apresentação e validação para se proceder a um processo de autenticação completo. A inscrição para permitir a leitura e o armazenamento nas bases de dados dos provedores de serviços de nuvem, em fases anteriores ao próprio controle de acesso, das características que identificam as *fingerprints* dos usuários. A apresentação para fornecer, no momento da autenticação, as características recém-coletadas no aparelho móvel para o ambiente de nuvem. A validação para permitir a comparação das características coletadas com as armazenadas na nuvem e, uma vez que haja congruência entre elas, viabilizar a autorização para o acesso aos recursos.

Cabe salientar, nesse contexto, que as capacidades de armazenamento e processamento disponíveis na nuvem são muito mais adequadas para a realização dessas comparações que as dos equipamentos móveis, e isso é uma das nítidas vantagens da utilização dos recursos da MCC voltados para a segurança do seu próprio ambiente.

Apesar da larga utilização, essa forma de biometria apresenta algumas desvantagens em relação a outras técnicas desta mesma categoria: a necessidade de ação explícita do usuário (não é transparente para o usuário) e a não pertinência de utilização para viabilizar uma autenticação contínua.

3.3.2- Autenticação baseada em reconhecimento facial

Outro método bastante utilizado para a autenticação de seres humanos perante equipamentos ou sistemas é o reconhecimento facial. Esse esquema viabiliza a análise de uma imagem para identificar uma face humana e a identidade a ela associada [INDRAWAN, 2013]. Tal qual na forma anterior, esse método também conta com uma quantidade bastante grande de equipamentos que viabilizam a captura das características biométricas a serem utilizadas nos processos de autenticação. A Figura 3.8 indica alguns desses equipamentos.

Figura 3.8 –Modelos de equipamentos móveis com câmeras

Algo que diferencia esses mecanismos de captura de imagem dos leitores de *fingerprint* anteriormente citados é o fato de eles terem sido inseridos nos seus equipamentos móveis não com o propósito de viabilizar segurança, mas sim com finalidades recreativas e até artísticas. Essas câmeras, entretanto, além de poderem ser utilizadas para registrar momentos interessantes por meio de fotografias e vídeos, também podem ser utilizadas para capturar características biométricas relacionadas à face do usuário para que sejam utilizadas em processos de autenticação.

Também há uma série de métodos de autenticação propostos para viabilizar o controle de acesso aos serviços de MCC [YE, 2015] [DEEPAK, 2015] [INDRAWAN, 2013] [PAWLE, 2013] baseados nesse tipo de equipamento e que obedecem a uma dinâmica semelhante à explorada anteriormente (etapas de inscrição, apresentação e validação) havendo uma diferença nítida: a fase de apresentação é subdividida em duas - a detecção da face dentro da imagem capturada e o reconhecimento das características que serão utilizadas no processo de autenticação [STOJMENOVIC, 2012].

Uma das grandes vantagens desses métodos é a possibilidade de uso dos equipamentos de captura de imagem (tipicamente a câmera frontal que sempre está voltada para o usuário) ao mesmo tempo em que se realizam diversas tarefas com o equipamento móvel. Dessa forma o usuário não precisa interromper suas atividades para realizar uma ação de autenticação e esta pode passar a ser transparente para o ser humano. Vinculado a isso, considerando que as capturas de imagens podem permanecer sendo feitas após a autenticação inicial (inclusive sem a necessidade de interferência do indivíduo), é possível ser realizado um processo de autenticação contínua do ser humano durante todo o período no qual ele esteja à frente da câmera.

3.3.3- Autenticação baseada na verificação da voz

A verificação de voz é uma forma de autenticação que utiliza características biométricas extraídas da voz que são únicas para cada ser humano. Essas características, desde que não haja mudanças bruscas no aparelho vocal do indivíduo (cirurgias de boca ou garganta, por exemplo), tendem a permanecer inalteradas (ou com poucas alterações) após o indivíduo atingir sua maturidade vocal (fase pós adolescência), o que faz da voz um bom parâmetro a ser utilizado na autenticação de usuários perante sistemas complexos como os baseados em MCC.

Durante a etapa de inscrição em um sistema de autenticação, a voz do indivíduo é gravada e dela é retirada aquilo que é conhecido como a *voiceprint* que será armazenado para fins de utilização nas etapas seguintes.

A fase de apresentação, por sua vez, apresenta uma sistemática diferente dos dois métodos abordados anteriormente, pois nela a captura das características não é feita a partir de um conteúdo instantâneo (tanto a *fingerprint recognition* quanto a *face recognition* são baseadas na análise de um conteúdo estático, um *snapshot*), mas sim a partir de um áudio que possui certa duração e que, geralmente, é orientado pelo próprio sistema de autenticação que sugere um conjunto de dígitos numéricos ou um texto pré-definido a ser lido no momento da autenticação [GUNSON, 2011]. A fase seguinte, a validação, segue os mesmos princípios que nos casos anteriores.

Algo que favorece bastante a aplicação desse tipo de autenticação para os equipamentos móveis é o fato de praticamente todos os equipamentos possuírem algum tipo de dispositivo de captura de voz (todos os *smartphones* e quase todos os *tablets* e *laptops* possuem microfones), o que já viabiliza, por definição, a possibilidade de funcionamento de um sistema baseado em voz.

Cabe salientar que a utilização desse tipo de característica biométrica para prover identificação e autenticação de indivíduos é diferente da utilização da voz para identificar as palavras que estão sendo pronunciadas (autenticação do usuário X reconhecimento de texto). Apesar de ambas utilizarem a mesma fonte de informação e poderem funcionar concomitantemente, os mecanismos de análise são diferentes e têm objetivos totalmente distintos.

Um aspecto vantajoso desse tipo de verificação, particularmente no momento de uma conversação telefônica (ou via VOIP), é a possibilidade de viabilizar, tal qual a *face recognition*, a autenticação contínua do indivíduo. Entretanto, para que isso seja possível, a base de dados de características de voz utilizada na autenticação deve ser extensa o suficiente para que não seja necessária a leitura de textos pré-definidos.

3.3.4- Autenticação baseada na verificação de sinais cardíacos

Considerando a relevância desse tema para o presente trabalho, esse item será explorado com um nível de detalhamento maior no Capítulo 4-Fundamentos de Autenticação baseada em Eletrocardiografia e de Aprendizado de máquina.

3.4- ASPECTOS DE AUTENTICAÇÃO CONTÍNUA

3.4.1- Autenticações contínua X estática e intrusiva X não intrusiva

Da leitura dos itens anteriores, percebe-se nitidamente que grande parte dos métodos de autenticação convencionais são estáticos, ou seja, são executados apenas para a abertura de uma sessão segura (partes autenticadas) não havendo a ideia de manutenção constante da autenticação realizada originalmente [CALDERON, 2006]. Um dos principais problemas desse enfoque está vinculado à execução de sessões relativamente extensas. Nesses períodos de tempo, é possível que o usuário autenticado saia de perto do equipamento utilizado para sua autenticação (um computador pessoal, um *smartphone*, etc.) gerando uma brecha para um indivíduo mal intencionado. Na ausência do usuário legítimo, o intruso poderá, então, operar o equipamento fazendo-se passar pelo usuário que realizou o processo de autenticação inicial. Os danos oriundos dessa ação são de diversos tipos e poderão atingir o usuário legítimo e a própria organização para a qual o usuário autenticou-se originalmente.

Nesse contexto, os processos de autenticação contínua mostram-se alternativas interessantes e viáveis para substituírem os processos estáticos. O enfoque contínuo implica na realização de tarefas de autenticação distribuídas por toda a duração da sessão segura (partes autenticadas). Após a execução do processo de autenticação inicial, o mecanismo responsável pela autenticação contínua passa a checar constantemente se o

usuário que está realizando as transações na sessão aberta é o mesmo que realizou a autenticação inicial [KLOSTERMAN, 2000].

As tarefas de autenticação geralmente são realizadas em instantes determinados a partir de períodos de tempo previamente definidos ou impostos por determinados tipos e quantidades de ações executadas durante a sessão.

É comum, por exemplo, que sejam executadas tarefas de autenticação a cada 30 segundos (ou período adequado à situação considerada). Note que a definição do período adequado é fundamental para manter a usabilidade do sistema. Caso o período seja muito pequeno (segundo a segundo, por exemplo) o processamento dedicado à atividade de autenticação poderá inviabilizar a execução de outras atividades. De maneira contrária, caso o período seja muito grande (hora a hora, por exemplo), a segurança do processo de autenticação poderá estar comprometida.

De outra forma, é possível que tarefas de autenticação sejam acionadas a partir do instante em que o usuário tenha realizado dez ações quaisquer (ou outra quantidade adequada ao caso analisado) após a última tarefa de autenticação.

Em outra situação, as tarefas podem ser acionadas a partir do momento em que determinados tipos de ações são executadas (geralmente ações críticas que demandam um nível de segurança maior que o usual).

Apesar de a autenticação contínua ser observada em todos os casos anteriores, percebe-se que, nesses casos específicos, os procedimentos de autenticação são realizados de forma discreta e em momentos específicos determinados por regras definidas previamente. Para essas situações é possível, inclusive, utilizar-se processos de autenticação clássicos (usuário e senha, certificados digitais, etc.) que serão disparados frequentemente durante a sessão.

Esse último enfoque é considerado intrusivo e é relativamente incômodo para o usuário tendo em vista que periodicamente ele deverá interromper suas atividades normais para realizar, de forma explícita, um processo de autenticação.

Uma opção interessante para minimizar esse problema é utilizar processos de autenticação não intrusivos que não demandem a interrupção das atividades normais de

operação do sistema (biometria, por exemplo) e funcionem de forma transparente para o usuário.

De forma mais sofisticada, é interessante misturar os dois tipos de mecanismos comentados (processos clássicos e biometria) aplicando a forma não intrusiva para o usuário na maior parte do tempo e a forma intrusiva (que demanda a interrupção da atividade normal do usuário) nos momentos em que a forma anterior não for capaz de viabilizar a autenticação desejada.

3.4.2- Níveis de autenticação contínua

Analisando os problemas da forma e do motivo de acionamento das tarefas de autenticação comentadas anteriormente, [CALDERON, 2006] propõe quatro níveis distintos de autenticação contínua, como visto a seguir:

Nível 1 – Autenticação do usuário

Nível no qual o usuário tem sua identidade verificada no início da sessão e, a partir de então, são realizadas tarefas de autenticação periódicas que mantêm o estado de "autenticado" do usuário por toda a duração da sessão. Para que as atividades normais do usuário não sejam interrompidas, devem ser utilizados, preferencialmente, métodos de autenticação não intrusivos e transparentes para o usuário. Entretanto, quando esses métodos não forem suficientes para atestar a autenticidade do usuário, é possível que seja utilizado um método intrusivo que gerará uma interrupção na utilização normal do sistema considerado.

Um exemplo simples de autenticação deste nível pode ser observado quando um usuário se autentica explicitamente para um sistema e passa a utilizá-lo normalmente, mas em um determinado momento precisa ausentar-se por alguns instantes. Quando o usuário regressa, após um determinado intervalo de inatividade na operação do sistema, este solicita uma reautenticação explícita para validar novamente a sua identidade. Nesse caso, a mera utilização normal do sistema, apesar de não ser um método convencional de autenticação, representa o procedimento não intrusivo indicado anteriormente. Um aspecto inconveniente observado neste exemplo é o fato do sistema só requerer um processo de reautenticação explícito caso ele permaneça inativo por um determinado intervalo de

tempo. Dessa forma, caso um usuário mal intencionado passe a utilizar o sistema tão logo o usuário autêntico ausente-se, o sistema não será capaz de deter essa falha de segurança.

Um exemplo mais sofisticado de processo do nível 1 considera a utilização de um mecanismo mais confiável de autenticação não intrusiva (um *token* conectado permanentemente ao sistema ou algum tipo de esquema de reconhecimento facial, por exemplo) durante o período de operação normal do sistema. Nesse caso específico, a deficiência observada no exemplo anterior não existirá.

Nível 2 – Autenticação usuário-recurso

Nessa situação, a autenticação genérica do usuário tratada no nível 1 permanece sendo realizada e, além disso, caso o usuário deseje acessar determinados recursos mais sensíveis do sistema, este solicitará uma reautenticação explícita. Note que essa reautenticação ocorrerá mesmo que o usuário esteja sendo autenticado continuamente pelo processo de autenticação não intrusiva comentado anteriormente.

Esse nível de autenticação tem o objetivo de evitar que um usuário autêntico acesse indevidamente recursos não permitidos para seu perfil (conjunto de privilégios), ou seja, mesmo sendo reconhecido como lícito perante o sistema, o usuário precisa possuir privilégios adequados e realizar autenticações específicas para poder acessar os recursos mais protegidos.

Nível 3 – Autenticação usuário-recurso-sistema

Nos dois níveis anteriores, o usuário considerado acessa o sistema por meio de equipamentos que fazem parte da rede de comunicações da própria organização detentora do sistema. Entretanto nem sempre isso é possível. Nesse contexto, o nível 3 viabiliza, além dos aspectos considerados nos níveis anteriores, a autenticação do usuário no caso em que este utiliza equipamentos remotos, posicionados em ambientes diferentes do local da própria organização e que, por esse motivo, não são controlados por ela.

Nível 4 – Autenticação usuário-recurso-sistema-transação

Há situações onde, além dos recursos serem sensíveis, as transações também são críticas. Operações que manipulem dados financeiros de uma organização, por exemplo,

não representam acessos a recursos, mas sim, a execução de transações críticas. Para esses casos, o nível 4 de autenticação contínua também impõe a execução de um processo explícito de autenticação, tal qual visto no nível 2 para o caso do acesso aos recursos restritos.

Dessa forma, no nível 4 serão impostas as autenticações tratadas nos níveis anteriores e as autenticações específicas para controlar o acesso a transações críticas que podem ser viabilizadas por meio do sistema operado.

Como pode ser percebido, este nível de autenticação fornece o mais alto grau de proteção verificando continuamente a autenticidade da identidade do usuário, quer ele esteja operando o sistema a partir do próprio ambiente da organização ou fora deste, além de considerar as necessidades específicas de autenticação para acesso a recursos sensíveis e a transações críticas.

3.5- AUTENTICAÇÃO MULTIMODAL

A maior parte dos sistemas de segurança envolvidos com o problema de controle de acesso (a sistemas, a instalações, a recursos específicos, etc.) utiliza uma ou mais técnicas que permitem validar a autenticidade da identidade do usuário, autorizando ou não o seu acesso [IKEHARA, 2010]. Nesse contexto, quanto maior a quantidade de processos de autenticação confiáveis utilizados para verificar a autenticidade, maior será a confiabilidade do sistema de controle de acesso.

É comum serem encontradas em organizações preocupadas com esse assunto, por exemplo, instalações e outros recursos que somente podem ser acessados a partir do momento em que o indivíduo prova sua identidade por meio de mais de um método de autenticação (cartão de acesso e senha, senha e reconhecimento de *fingerprint*, etc.). Esse tipo de processo que pressupõe a utilização conjugada de diversos métodos é conhecido como autenticação multimodal e geralmente também funciona vinculado ao conceito de autenticação contínua.

Dessa forma, após o processo de autenticação inicial (tipicamente fundamentado em métodos clássicos de autenticação e/ou em biometria fisiológica), não apenas um, mas vários métodos passam a funcionar paralelamente e constantemente para viabilizar a

autenticação contínua desejada. Apesar de claramente demandar maiores recursos de processamento do ambiente onde o sistema de autenticação funciona, essa forma apresenta a grande vantagem de permitir que o processo de autenticação não seja interrompido, mesmo que um dos métodos não gere os resultados esperados.

Em casos extremos, caso todos os métodos envolvidos na autenticação contínua e multimodal falhem (o que é pouco provável), pode ser, ainda, imposto ao usuário um método estático intrusivo que servirá de base para a autenticação.

Sistemas de autenticação multimodal geralmente utilizam processos não intrusivos. Isto é necessário porque os processos devem funcionar de forma paralela e constante e, caso eles dependam da intervenção do usuário, este provavelmente não terá tempo, sequer, para realizar a operação normal do sistema cujo acesso está sendo controlado. Nesse contexto percebe-se a importância da biometria. Grande parte dos processos baseados em biometria, pelo fato de não serem intrusivos, podem ser utilizados paralelamente sem que o usuário perceba, viabilizando a existência de sistemas biométricos multimodais que reconhecem características fisiológicas e/ou comportamentais do usuário concedendo ou inabilitando seu acesso aos recursos desejados [AZZINI, 2008], [ANAND, 2010].

3.6- CONCLUSÃO

O presente capítulo tem o propósito de apresentar fundamentos teóricos sobre autenticação e, de forma mais específica, biometria, que pode ser considerada uma das principais maneiras de viabilizar a autenticação de seres humanos.

Aqui são apresentados os processos convencionais de autenticação mais utilizados no mundo e alguns aspectos gerais de biometria. Após isso, são detalhados, de forma abrangente, vários métodos de autenticação baseados em biometria, dentre os quais pode-se citar os fundamentados em impressão digital e da mão, reconhecimento facial, verificação de voz e verificação de sinais cardíacos (um dos principais focos desta tese). Por fim, são explorados conceitos de autenticação contínua e multifatorial.

Como conclusão parcial, é possível considerar que os temas abordados apresentam várias técnicas importantes para viabilizar a autenticação de usuários de sistemas, quer seja por meio de algo que eles possuam, saibam ou faça parte deles.

4- FUNDAMENTOS DE AUTENTICAÇÃO BASEADA EM ELETROCARDIOGRAFIA E DE APRENDIZADO DE MÁQUINA

Para que seja possível compreender a lógica desse tipo de autenticação, são necessários alguns fundamentos teóricos que são abordados neste capítulo, dentre os quais se destacam: a anatomia e o sistema de condução do coração, a eletrocardiografia e sua aplicação na biometria, e noções de aprendizado de máquina.

A Tabela 4.1 apresenta as variáveis e constantes utilizadas neste capítulo.

4.1- ASPECTOS DO FUNCIONAMENTO DO CORAÇÃO

O coração, que em conjunto com os vasos sanguíneos e o sangue formam o sistema circulatório do corpo humano, é uma bomba muscular que faz o sangue circular por todo o corpo, de forma contínua, durante toda a vida de um indivíduo. Ele pulsa, considerando uma vida média de 70 anos de uma pessoa, aproximadamente 2,5 bilhões de vezes [TORTORA, 2014]. Por minuto e em seu estado basal (frequência cardíaca normal mínima), seu lado esquerdo bombeia aproximadamente 5 litros de sangue arterial - rico em oxigênio - para a maior parte dos vasos sanguíneos do corpo e seu lado direito empurra um volume semelhante de sangue venoso - rico em dióxido de carbono - para os pulmões a fim de que esse gás seja substituído por oxigênio.

4.1.1- Anatomia do coração

O coração de uma pessoa adulta assemelha-se a um cone com, aproximadamente, 12 cm de altura e 9 cm de diâmetro em sua maior sessão transversal, possuindo quase o volume de uma mão fechada e tendo uma massa aproximada de 300 gramas.

Ele está posicionado próximo ao centro da cavidade torácica, numa região conhecida como mediastino, e possui aproximadamente 70% de sua massa na porção esquerda dessa cavidade, indicando uma ligeira assimetria em relação ao eixo central do corpo humano. A Figura 4.2 ilustra o posicionamento e a forma do órgão.

Tabela 4.1 - Variáveis utilizadas neste capítulo

Variável/ constante/ função	**Descrição**
tp	Número de verdadeiros positivos
tn	Número de verdadeiros negativos
fp	Número de falsos positivos
fn	Número de falsos negativos
Acc	Valor da acurácia
Pre	Valor da precisão
Sen	Valor da sensibilidade
$F1s$	Valor do escore F1
l	Rótulo de um classificador
$x_1, x_2,..., x_m$	Variáveis que representam pontos discretos de um domínio de análise de um probl. de classificação (exemplos submetidos a um treinamento de classificador)
$C_0, C_1,..., C_m$	Classes possíveis de uma classificação também chamadas de rótulos
$y_1,y_2,..., y_m$	Variáveis que assumem os valores das Classes possíveis de uma classificação
$sign( )$	Função sinal que fornece o sinal do resultado de $( )$ ou 0 caso o resultado seja nulo
w	Vetor perpendicular ao hiperplano ótimo de separação de classes em uma SVC em um espaço n-dimensional
$h_1, h_0, h_{-1}, h_1', h_0', h_{-1}'$	Hiperplanos de um espaço n-dimensional
w	Vetor perpendicular ao hiperplano ótimo no espaço de n dimensões
b	Valor escalar ou constante
K	Função de *kernel*
ϕ	função que transforma um vetor x de um espaço n-dimensional em um vetor equivalente em espaços de dimensões mais altas (até infinitas)
x'_j	Variáveis não rotuladas (cuja classe correspondente é desconhecida)
$\boldsymbol{x}_i^T$	Vetor de variáveis de treinamento em um domínio transformado
γ, r e d	Parâmetros do *kernel* de uma SVM
D	Distribuição amostral considerada no treinamento de um classificador
$Class$	Classificador específico
lo_{exp}	Função de perda exponencial
f	Função que retorna o rótulo de x
$E_{\sim D}[f( )]$	Esperança matemática de uma função f
$I( )$	Função que cujo resultado é 1 se $( )$ é verdadeiro e 0 caso contrário
T	Número total de iterações de um treinamento ou quantidade de classificadores fracos que fazem parte de um processo de *boosting*
t	Uma iteração de um treinamento ou índice que um dos classificadores que fazem parte de um processo de *boosting*
α	
θ	Margem a ser atingida no processo de minimização de erro
M	Espaço n-dimensional
p	Ponto de M
S	Conjunto de pontos p
q	Ponto de M que serve de referência para os demais

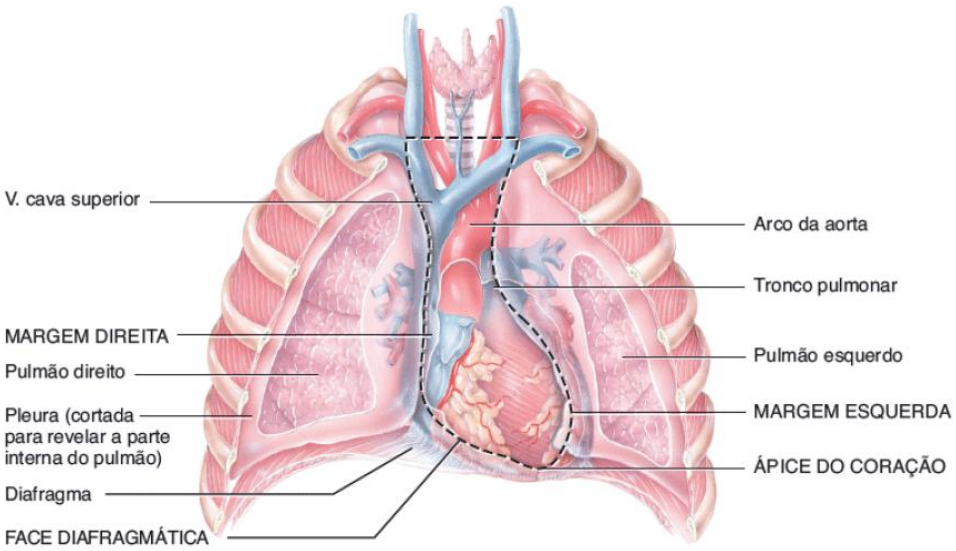

Figura 4.1 –Posição e forma do coração
[TORTORA, 2014].

O coração é envolvido por uma membrana conhecida como pericárdio que o mantém posicionado junto ao mediastino, protege e possibilita liberdade suficiente para que o processo de pulsação ocorra normalmente.

As paredes do órgão são divididas em três camadas:

- Epicárdio - porção mais externa que confere uma textura lisa à superfície do coração e contém os principais vasos cardíacos, sendo rica em gordura (proporcionalmente à gordura corporal geral do indivíduo);
- Miocárdio - porção média que ocupa, aproximadamente, 95% da massa total das paredes e é composta pelo tecido muscular responsável pela pulsação;
- Endocárdio - porção mais interna que fornece um revestimento liso para as câmaras do coração, facilitando o fluxo sanguíneo, e possui as válvulas cardíacas responsáveis pelo controle desse fluxo.

O coração possui 4 câmaras e 4 válvulas responsáveis por manter separadas as porções de sangue que chegam e saem do órgão. Dois átrios (um direito e outro esquerdo) localizados na parte superior que recebem o sangue vindo do resto do corpo e dois ventrículos (um direito e outro esquerdo) na parte inferior que despacham o sangue para os vasos que o distribuirão pelo corpo. A Figura 4.2 ilustra as câmaras do coração.

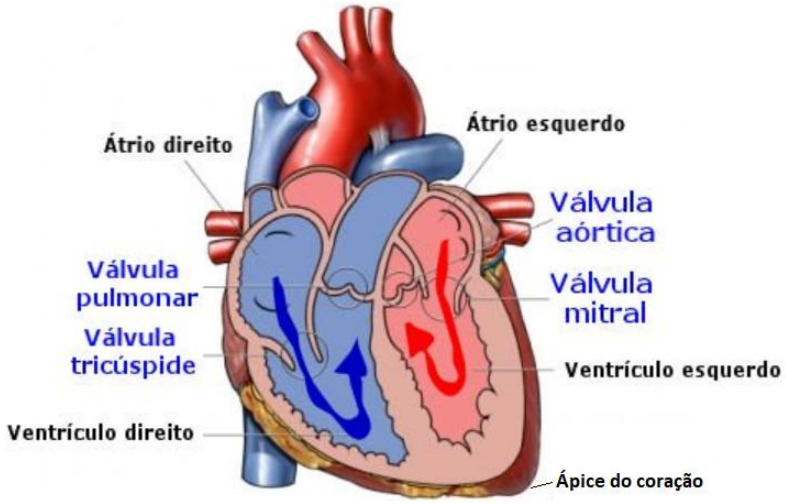

Figura 4.2 –Câmaras e válvulas do coração
[MD SAUDE, 2019].

O sangue venoso (rico em dióxido de carbono) chega à porção direita do coração (azul na figura) trazido pela veia Cava superior e é armazenado, inicialmente, no átrio. A válvula Tricúspide é responsável por separar o átrio direito do ventrículo direito (válvula atrioventricular). Acionada por contrações musculares, essa válvula abre e o sangue do átrio é empurrado para o ventrículo, onde fica bloqueado, pois a válvula seguinte, a Pulmonar (válvula semilunar), fecha quando a anterior, a Tricúspide, abre. Por meio de nova contração, a válvula Tricúspide fecha e a Pulmonar abre jogando o sangue para as artérias Pulmonares (esquerda e direita) que levarão os fluxos até os pulmões.

De forma análoga ao sangue venoso, o sangue arterial (rico em oxigênio) chega à porção esquerda do coração (vermelho na figura) trazido pelas veias Pulmonares (esquerda e direita) e é armazenado, inicialmente, no átrio. A válvula Mitral é responsável por separar o átrio esquerdo do ventrículo esquerdo (válvula atrioventricular). Acionada por contrações musculares, essa válvula abre e o sangue do átrio é empurrado para o ventrículo, onde fica bloqueado, pois a válvula seguinte, a Aórtica (válvula semilunar), fecha quando a anterior, a Mitral, abre. Por meio de nova contração, a válvula Mitral fecha e a Aórtica abre jogando o sangue para a veia Aorta que levará o fluxo ao resto do corpo.

Como pode ser visto, todo o fluxo sanguíneo que passa pelo coração é provocado por suas contrações musculares. Essas contrações são, por outro lado, desencadeadas por meio de pequenos pulsos elétricos, algo que será detalhado à frente.

4.1.2- O sistema de condução do coração

Resumindo o que foi descrito até aqui sobre o coração, esse órgão funciona como uma grande bomba que puxa e impulsiona sangue para todo o corpo. Para que isso seja possível, a musculatura do coração deve trabalhar continuamente e suas fibras devem ser estimuladas por meio de pulsos elétricos. A origem desses pulsos elétricos está relacionada a uma rede de fibras musculares chamadas de "fibras autorrítmicas" que produzem continuamente potenciais de ação que geram as contrações cardíacas.

Essas fibras existem em quantidades bastante pequenas se comparadas aos demais tipos de fibras musculares cardíacas. Elas têm duas funções fundamentais:

- Funcionar como um marca-passo definindo o ritmo cardíaco;
- Compor o chamado sistema de condução do coração, uma rede de fibras específicas que fornecem um caminho para que cada ciclo de excitação cardíaca se propague adequadamente pelo coração.

O sistema de condução do coração permite que as diversas partes do coração funcionem de forma coordenada gerando as contrações musculares necessárias para o fluxo sanguíneo.

A excitação elétrica do coração geralmente começa no átrio direito (especificamente no nó sinoatrial - SA) [TORTORA, 2014] e se espalha pelos vários locais do coração por meio do sistema de condução, desencadeando o ciclo cardíaco. O nó sinoatrial, pelo fato de possuir uma despolarização espontânea muito rápida (a cada 0,6 s), funciona como um marca-passo natural do coração. Isso ocorre porque as outras fibras autorrítmicas, sendo mais lentas do que as do SA, sofrem despolarizações estimuladas pela despolarização do SA antes de iniciarem suas próprias despolarizações autônomas.

O funcionamento elétrico do coração, apesar de ser estimulado, em muito, pelas fibras do próprio órgão, recebe estímulos externos que interferem no ciclo cardíaco. Esses estímulos podem ter origens nos pulsos elétricos do Sistema Nervoso Autônomo - SNA ou na presença de hormônios que trafegam pelo corpo por meio da corrente sanguínea.

A interação dos estímulos internos (fibras autorrítmicas) e externos (SNA e hormônios) gera uma série de pulsos elétricos que podem ser mapeados no ciclo cardíaco.

Um potencial de ação observado no coração ocorre em três fases: despolarização (descarga elétrica que gera um aumento do potencial elétrico das células), platô (após a descarga elétrica, o potencial é mantido inalterado por alguns instantes) e repolarização (recuperação ou retorno ao potencial do repouso). Essas três fases ocorrem em momentos distintos nas diferentes áreas do coração, o que gera uma série de diferenças de potencial que podem ser medidas e representadas por meio de eletrocardiogramas.

4.1.3- Eletrocardiografia

As várias alterações de potencial causadas pelas diferentes interações entre as despolarizações, os platôs e as repolarizações das diferentes fibras que compõem o sistema de condução do coração podem ser capturadas e representadas por meio de um artefato bastante útil, o eletrocardiograma - ECG.

O ECG é um método que se propõe a realizar medições e registros de sinais elétricos emitidos pelo coração sem que haja necessidade de atingir a superfície desse órgão, pois a medição é feita por meio da justaposição de eletrodos diretamente na pele do indivíduo [SADDIK, 2017]. Esse método é considerado não invasivo e mede as irregularidades observadas no funcionamento do coração que representam os ciclos de contração e relaxamento dos átrios e dos ventrículos (cavidades existentes no coração) [SINGH, 2012] e que estão intimamente ligados aos formatos que esse órgão assume nas pessoas.

As principais características de um ECG estão relacionadas aos 5 pontos fiduciais principais (P, Q, R, S, T) característicos da onda cardíaca [PATRO, 2017] e que representam seus picos (P, R, T) e vales (Q, S), como pode ser observado na Figura 4.3.

O primeiro pulso, a "onda P", está relacionado à despolarização do nó sinoatrial mencionado anteriormente, sendo positiva (aumento do potencial elétrico) e tendo uma duração inferior a 120 ms. Em seguida, as alterações nos potenciais elétricos derivadas da atividade ventricular geram o chamado "complexo QRS", cuja duração varia entre 70 e 110 ms [REZGUI 2016]. Por último, a repolarização ventricular gera a "onda T" que tem uma

duração aproximada de 300 ms. É importante salientar que todos os intervalos indicados são obtidos do funcionamento de um coração saudável.

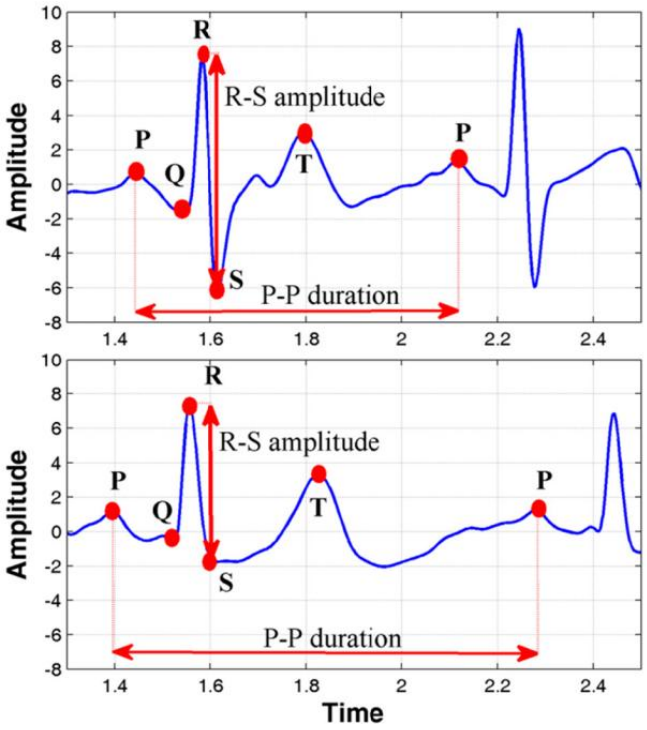

Figura 4.3 – Representação de ECG de dois indivíduos diferentes
[ISLAM, 2017]

As diferenças existentes, de indivíduo para indivíduo, entre os posicionamentos relativos dos cinco pontos característicos indicados, considerando tempo e amplitude, são causadas por distinções de aspectos fisiológicos como, por exemplo, posição do coração em relação a outros órgãos, tamanho absoluto e condições físicas da pessoa.

Na Figura 4.3 são indicados, como exemplos, os parâmetros "R-S *amplitude*", que representa a distância medida entre os pontos R e S no domínio da amplitude, e "P-P *duration*", que indica o intervalo de tempo existente entre dois pontos P consecutivos. É possível observar, mesmo a olho nu, que há diferenças nítidas entre esses parâmetros obtidos de dois indivíduos diferentes, o que sinaliza a possibilidade de utilização dessa característica como algo que identifique as pessoas envolvidas.

Em análises mais completas ([FALCONI, 2016] e [REZGUI, 2016], por exemplo), a quantidade de parâmetros (distâncias em amplitude e tempo, ângulos relativos entre retas

que ligam pontos fiduciais, etc.) é bem maior que no exemplo da figura anterior e permite uma maior precisão na identificação dos indivíduos.

Apesar das diferenças apontadas entre os parâmetros do ECG analisados poderem indicar medições extraídas de pessoas diferentes, um mesmo indivíduo também possui ECG com formatos diferentes. A Figura 4.4 mostra várias ondas sobrepostas extraídas do ECG de um mesmo indivíduo.

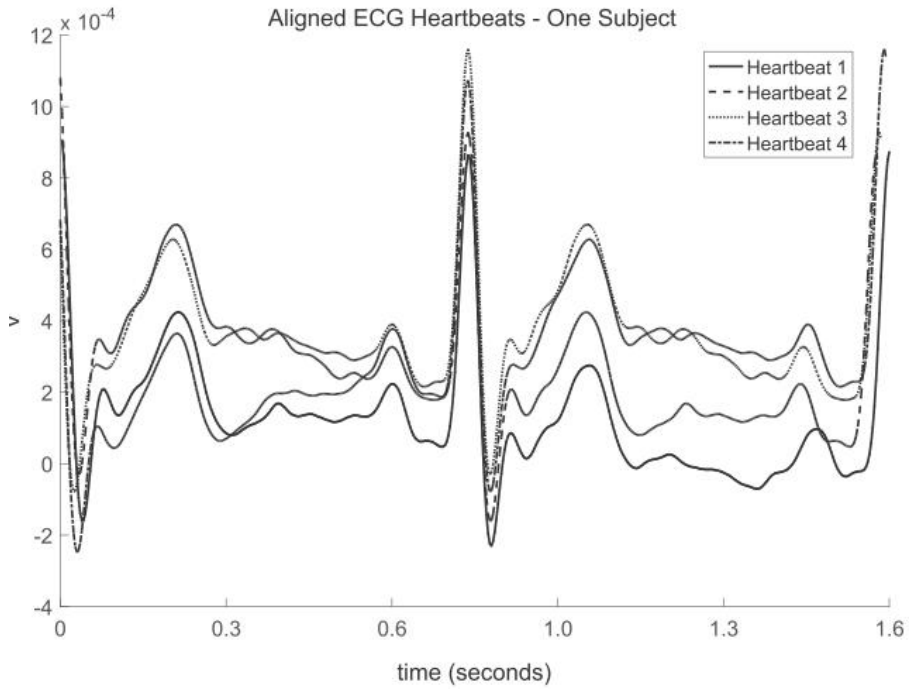

Figura 4.4 – Alinhamento de ondas de ECG de um mesmo indivíduo [FALCONI, 2018]

Na figura percebe-se que, apesar das ondas cardíacas terem sido extraídas de um mesmo indivíduo, há várias diferenças entre as mesmas, o que pode ensejar análises equivocadas em relação à identificação da pessoa que as gerou. Por esse motivo, a utilização de ECG para a identificação de pessoas é algo que deve ser precedida pela aplicação de técnicas de processamento de sinais para, por exemplo, filtrar ruídos, e pela seleção adequada dos parâmetros que realmente não variarão entre eletrocardiogramas de um mesmo indivíduo.

4.2- BIOMETRIA BASEADA EM ELETROCARDIOGRAFIA

4.2.1- Aspectos gerais

A biometria baseada em sinais cardíacos fundamenta-se em dados extraídos de ECG que podem ser utilizados para a autenticação de indivíduos perante uma grande variedade de sistemas (inclusive envolvendo o uso de equipamentos móveis) [SADDIK, 2017]. O contato com locais pré-determinados da pele do indivíduo feito por eletrodos ou sensores óticos existentes em diversos equipamentos (sensores específicos, relógios, *smartphones*, etc.) permite a obtenção de dados de ECG que, após passarem por processamentos adequados, viabilizam a extração de parâmetros representativos de um indivíduo que podem ser comparados a parâmetros análogos armazenados em bases de dados de padrões, o que pode viabilizar não somente a identificação de um ser humano, mas também a sua autenticação.

Muito se tem feito para que a precisão dos resultados de identificação e/ou autenticação baseados em ECG seja cada vez maior, entretanto a ideia de que esse método ainda não é preciso o suficiente para esses fins ainda é algo indicado por alguns autores [SINGH, 2012]. Entretanto esses mesmos autores concordam que a utilização da biometria baseada em ECG, em conjunto com outros métodos de um sistema multimodal, é algo que aumenta sobremaneira a precisão e robustez desse sistema.

A autenticação baseada em ECG pode ser considerada não intrusiva e transparente para o usuário desde que sejam utilizados os sensores adequados (como visto à frente). Essas características indicam claras vantagens em relação ao método de autenticação baseado em impressão digital, o qual não é considerado transparente para o usuário. Além disso, essa forma de biometria apresenta vantagens de segurança em relação a outros métodos de autenticação biométrica fisiológica descritos anteriormente. Em relação à autenticação por meio da verificação da voz, esta é vulnerável a ataques simples baseados na reprodução de gravação da voz do usuário. Já no caso da autenticação baseada em reconhecimento facial, este método pode ser burlado com o simples uso de imagens (estáticas ou dinâmicas) previamente capturadas do usuário. Acerca desses dois últimos métodos, percebe-se que os mesmos podem ser burlados com a apresentação, para os seus sensores (microfone e câmera) de conteúdos simples (gravação de voz e imagens capturadas). Essa mesma situação não é observada para o método baseado em ECG (é mais

difícil apresentar para um dos sensores apresentados um padrão representativo de um ECG).

4.2.2- Sensores cardíacos

A Figura 4.5 ilustra uma série de dispositivos que podem ser utilizados para a captura de sinais cardíacos. Nessa figura, com exceção dos dois mais à direita, todos os equipamentos viabilizam uma utilização não intrusiva e transparente para o usuário.

Figura 4.5 –Modelos de sensores cardíacos comerciais

Nos exemplos indicados, há eletrodos que devem ser colocados em contato com a pele da região torácica (os cinco modelos mais à esquerda), sensores óticos que devem ser colocados no pulso (os três modelos centrais) e outros sobre os quais se deve colocar o dedo de forma semelhante àquilo que se faz com um leitor de impressões digitais (dois modelos à direta).

Pelo fato dos sensores não possuírem capacidade própria de processamento, os dados são extraídos (com maior ou menor precisão) em todos os casos indicados e transmitidos para processadores (existentes em relógios, *smartphones*, etc.) que, utilizando algoritmos apropriados, os processam gerando os parâmetros adequados às diversas finalidades (monitoramento da situação cardíaca, identificação e/ou autenticação de indivíduos, etc.).

Tendo percebido os diversos parâmetros que podem ser extraídos a partir do funcionamento elétrico do coração, é necessário, agora, indicar alguns métodos para classificar esses parâmetros de tal forma que eles possam ser utilizados em processos de autenticação.

4.3- APRENDIZADO DE MÁQUINA (*MACHINE LEARNING*)

4.3.1- Fundamentos

Aprendizado de máquina (do inglês, *Machine Learning* - ML) pode ser entendido como um ramo da computação no qual algoritmos tornam-se capazes de tomar decisões com base em aprendizados fundamentados em dados submetidos à análise desses algoritmos. A grande ideia na qual o ML se baseia é a apresentação de grandes quantidades de dados (quanto maior a quantidade, melhor tende a ser o aprendizado) para algoritmos que, tendo aprendido com esses dados, terão condições de fornecer soluções para problemas inferenciais complexos [SIMEONE, 2017].

A origem da expressão "Aprendizado de máquina" não é recente e remonta à época em que os primeiros computadores foram criados (década de 1950), entretanto o ML assumiu proporções bastante relevantes nos últimos anos com o aumento das capacidades computacionais disponíveis em larga escala (particularmente nas nuvens computacionais).

Um dos fundamentos do ML é a otimização matemática que envolve a computação numérica voltada para a tomada de decisões baseadas em dados inéditos (que não foram utilizados durante o processo de aprendizado) [BOTTOU, 2016].

Dentre as áreas onde o ML tem sido bastante utilizado, destaca-se a de reconhecimento de padrões (do inglês, *Pattern Recognition* - PR), que se refere à descoberta automática de características (do inglês, *features*) existentes em uma determinada massa de dados com o objetivo de fundamentar processos de tomada de decisão e de predição, dentre outros. É importante frisar que essa massa de dados pode ter origens de diversos formatos (arquivos de imagem e de som, por exemplo), o que confere à área de reconhecimento de padrões uma versatilidade bastante interessante.

Uma das grandes motivações da área de ML é substituir o ser humano (ou pelo menos apoiá-lo) em trabalhos complexos que envolvem a definição e a utilização de

modelos que demandam a participação de uma gama grande de pessoas. A pesquisa e o desenvolvimento de algoritmos de reconhecimento facial, de reconhecimento de voz e de várias outras características biométricas são bons exemplos de trabalhos de computação que podem ser melhorados ou otimizados com a utilização de ML.

Este trabalho, especificamente, utiliza ML para reconhecer padrões em sinais de eletrocardiografia.

Modos de Aprendizado de Máquina

Em geral, é possível dividir os modos de ML em três tipos [SIMEONE, 2017]:

- Aprendizado supervisionado (do inglês, *Supervised Learning* - SL):
  Este modo fundamenta-se no aprendizado (ou treinamento) do algoritmo de ML utilizando dados conhecidos (já existentes em uma base de dados ou algo do gênero) e seus rótulos (classes às quais esses dados pertencem). Esse treinamento permite a criação de padrões baseados nesses dados fornecidos e possibilita a classificação de dados inéditos com base nos padrões (ou no modelo preditor). Dessa forma, pode-se dizer que o aprendizado supervisionado gera um relacionamento entre dados conhecidos e dados inéditos [KOHAVI, 1998].
  Em [ZHOU, 2012b], indica-se que, caso o SL esteja baseado em rótulos categóricos (ou discretos e pertencentes a um conjunto finito), ele é chamado de classificação e, caso seja fundamentado em rótulos numéricos contínuos (números reais, por exemplo), será chamado de regressão.
  No processo de treinamento indicado, cada instância do conjunto de dados e rótulos é chamada de exemplo e, para a classificação binária (dois valores possíveis), é típico utilizar-se exemplos positivos (que contêm dados e rótulos vinculados, ou seja, o dado é da classe representada pelo rótulo) e negativos (que contêm dados e rótulos desvinculados, ou seja, o dado não é da classe representada pelo rótulo).
  Uma aplicação do SL voltada para a área de segurança da informação é a utilização de grandes bases de dados de artefatos maliciosos (do inglês, *malware*) para treinar algoritmos que serão aplicados, posteriormente, na detecção de ameaças a sistemas computacionais.
  O presente trabalho faz uso deste modo de ML (da classificação, especificamente) por meio do seu emprego no processo de autenticação de um indivíduo (ou usuário

de um sistema) baseada em seus sinais de ECG. A opção por esse modo de aprendizado reside no fato da existência prévia (anterior aos momentos de treinamento das máquinas) de dados rotulados de ECG ser uma premissa para o modelo de autenticação proposto.

- Aprendizado não supervisionado (do inglês, *Unsupervised Learning* - UL):
  Esse modo de ML, diferentemente do anterior, não utiliza treinamento baseado em rótulos. O grande propósito, neste caso, é descobrir alguma classe inerente aos dados desconhecendo uma classificação prévia desses [ZHOU, 2012b].
  O agrupamento (do inglês, *clustering*) é uma das grandes aplicações desse modo de ML. Nesse caso, a ideia central é agrupar exemplos que possuam similaridades [SIMEONE, 2017]. Com isso, esses exemplos passam a ser representados por meio de uma quantidade menor de classes. Em outros termos, antes desse tipo de treinamento ser realizado, cada exemplo é considerado uma classe específica e, após o treinamento, vários exemplos agrupam-se, os grupos passam a ser representados por classes e, dessa forma, as quantidades de classes diminuem em relação à quantidade original.
  O aprendizado não supervisionado tem várias aplicações e destacam-se as que buscam anomalias em grandes massas de dados de diversas origens (financeiras, climatológicas, comportamentais, dentre outras).
- Aprendizado por reforço (do inglês, *Reinforcement Learning* - RL):
  Essa é a forma de ML que mais se assemelha aos aprendizados cotidianos pelos quais os seres humanos (e alguns animais irracionais) passam. Nesse contexto, cabe salientar que, na psicologia (especificamente no behaviorismo), um reforço é algo que torna um determinado comportamento mais provável de acontecer. Em oposição, uma punição tem efeito contrário, ou seja, diminui a probabilidade de ocorrência de certo comportamento. Analogamente, esse modo de ML baseia-se em reforços e punições que são impostos a um processo computacional a partir dos resultados de suas ações [SIMEONE, 2017].
- Dentre as várias aplicações de aprendizado por reforço, pode-se citar a área de jogos sobre os quais podem ser aplicados algoritmos computacionais. No xadrez, por exemplo, a utilização de RL é feita por meio da atribuição, a cada jogada desencadeada por um algoritmo de RL (chamado de agente), de um valor que pode representar reforços (maiores ou menores) ou punições (análogas aos reforços).

Com base nesses valores, o algoritmo deve, então, maximizar os seus resultados.

Classificação

A classificação é considerada o principal tipo de aprendizado de máquina [SIMEONE, 2017], possuindo aplicações em diversas áreas distintas como, por exemplo, a detecção de *spam* em bases de correio eletrônico e o diagnóstico de problemas de saúde.

No contexto da classificação, uma máquina que aprende é chamada simplesmente de classificador. Neste problema de SL, um rótulo l pode assumir, como indicado anteriormente, um número finito de valores discretos e, em uma classificação binária, cada ponto x de um domínio de análise pertence a uma das duas classes existentes (C_0 ou C_1) que são identificadas pelo rótulo l de acordo com ($x \in C_0\ se\ l = 0\ ou\ l = -1$) e ($x \in C_1\ se\ l = 1$).

É importante salientar que, em classificações que possuam maiores quantidades de classes (C_0, C_1, ..., Cm), não é recomendável que l de vários valores identifiquem a mesma classe.

A Figura 4.6 apresenta um exemplo de um problema de classificação binária no qual, a partir de um conjunto de treinamento D formado pelos exemplos x_1, x_2, ..., x_m , deseja-se classificar um novo exemplo $x \notin D$ em uma das duas classes existentes C_0 e C_1.

Como indicado anteriormente, este trabalho utiliza o aprendizado supervisionado (a classificação, especificamente) para viabilizar a autenticação de seres humanos baseada em seus sinais de ECG. Algumas *features* dos ECG dos usuários, como é detalhado à frente, neste trabalho, são utilizadas para realizar treinamentos de classificadores em momentos anteriores ao processo de autenticação e, durante a autenticação propriamente dita, novas *features* capturadas de ECG dos indivíduos são submetidas aos classificadores para que seja possível a verificação de suas classes e a conseguinte autenticação ou rejeição do usuário.

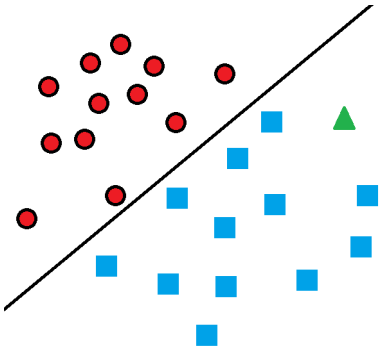

Figura 4.6 - Exemplo de classificação binária
Círculos vermelhos ∈ C_0, quadrados azuis ∈ C_1 e o triângulo verde é um novo exemplo que precisa ser classificado.

São 4 os classificadores escolhidos para serem avaliados neste trabalho:

- Máquinas de vetores de suporte (do inglês, *Support Vector Machines*- SVM);
- *AdaBoost*;
- *Robust Boost*;
- Buscador pelo Vizinho mais próximo (do inglês, *Nearest Neighbor Searcher* - NNS).

A escolha desses classificadores baseou-se nas pesquisas bibliográficas feitas para embasar este trabalho e que apontaram para algumas situações:

- Os quatro classificadores têm um desempenho comparativamente melhor que outras técnicas de ML quando são empregadas quantidades pequenas de exemplos para os treinamentos;
- Há trabalhos que usam NNS e SVM para classificar dados de ECG e que podem ser utilizados em comparação com o atual trabalho;
- Não foram encontradas, até o presente momento, referências que utilizem os algoritmos *AdaBoost* e *Robust Boost* para resolver o problema de classificação que é alvo desta tese. Por outro lado, esses algoritmos têm encontrados bons desempenhos em tarefas de classificação similares ao problema deste trabalho.

4.3.2- Máquinas de Vetores de Suporte

Máquinas de vetores de suporte são classificadores binários que mapeiam (originalmente de maneira linear, mas podendo ser adaptada para maneiras não lineares)

vetores de dados de entrada em um espaço *n*-dimensional de *features* [CORTES, 1995]. As SVM são métodos de SL que podem ser usados tanto para a classificação quanto para a regressão. Esses métodos fundamentam-se em treinamentos baseados em um conjunto de exemplos rotulados a partir do qual, com a utilização do algoritmo de treinamento, é criado um modelo preditor da classe de novos exemplos.

Esse tipo de aprendizado de máquina é utilizado em várias áreas, da informática à medicina, tratando problemas de PR para a filtragem de conteúdos de pacotes de dados que trafegam pelas redes de computadores, para o reconhecimento facial e de várias outras características biométricas, para a análise da escrita à mão e o reconhecimento de caracteres, dentre vários outros exemplos.

A base teórica sobre a qual as SVM apoiam-se deriva da teoria de Vapnik-Chervonenkis [VAPNIK, 2004] que tenta, considerando um enfoque de aprendizado computacional, explicar o processo de aprendizagem sob um ponto de vista estatístico.

Hiperplanos

Sob uma ótica geométrica, a classificação por vetores de suporte (do inglês, S*upport Vector Classification* - SVC) é aquela na qual um algoritmo busca por uma superfície *n*-dimensional ótima que separe, de maneira equidistante, as classes do problema em análise. Essa superfície é tipicamente conhecida como hiperplano [BHAVSAR, 2012]. A SVC é, inicialmente, descrita para problemas de separação linear. Em situações mais sofisticadas nas quais a solução linear não seja aplicável, podem ser utilizadas soluções não lineares baseadas em funções de núcleo (do inglês, *kernel*).

A Figura 4.7 apresenta alguns hiperplanos em um espaço *n*-dimensional (para a representação gráfica desse exemplo, $n = 2$). São utilizados m pontos rotulados (ou exemplos de treinamento) do tipo $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ nos quais os valores de y representam os valores das classes (para o caso, $y = -1$ ou $y = 1$). Esses hiperplanos representam o classificador linear dado por

$$y = sign(w^T \cdot x \pm b),$$

(4.1)

onde *sign* é a função *sinal* que fornece o sinal do resultado ou 0 caso o resultado seja nulo, b é um valor escalar e w^T é um vetor perpendicular ao hiperplano ótimo no espaço de n dimensões.

Os hiperplanos limítrofes das classes são dados, então, por $(w \cdot x - b = 1)$ e $(w \cdot x - b = -1)$, e o hiperplano ótimo para a separação das classes é dado por $(w \cdot x - b = 0)$,

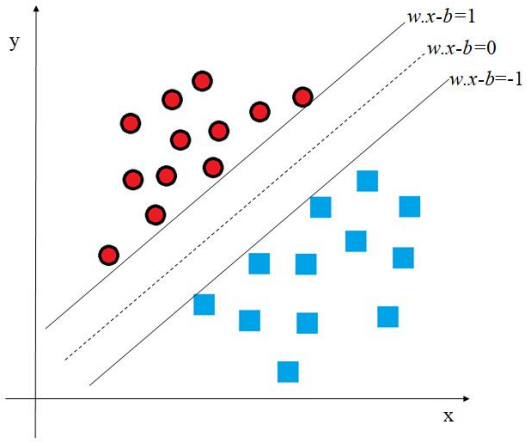

Figura 4.7 - Hiperplanos de uma SVM treinada para uma classificação binária

Como pode ser percebido, o hiperplano de separação ótima localiza-se em uma posição equidistante em relação aos dois outros hiperplanos. Na Figura 4.7, o hiperplano ótimo é representado pela linha pontilhada, enquanto os outros dois que sinalizam os limites das classes são representados por linhas contínuas.

Margens

É importante frisar que, sob a ótica geométrica, quanto maior for a distância (ou margem) entre os hiperplanos limítrofes das classes, menor será o erro de generalização do classificador. Dessa forma, durante o treinamento busca-se maximizar a margem.

A Figura 4.8 apresenta dois exemplos nos quais, apesar de existirem hiperplanos ótimos e limítrofes das classes, as margens têm tamanhos distintos. Entre h_1 e h_{-1}, à esquerda, a margem é maior que entre h'_1 e h'_{-1}, à direita.

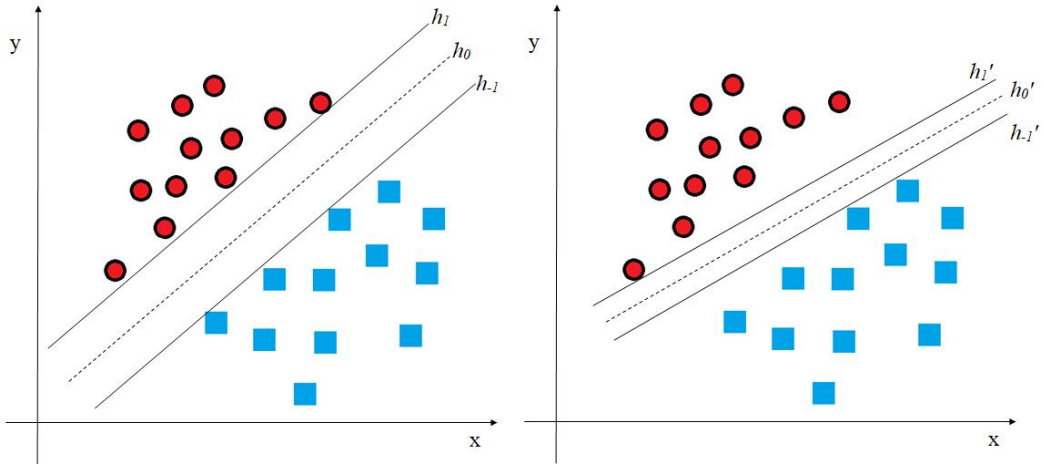

Figura 4.8 - Margens variáveis entre hiperplanos de uma SVM.

Os vetores de suporte são, nesse contexto, formados pelos dados de treinamento do espaço *n*-dimensional (exemplos) que definem a máxima margem de separação entre as classes.

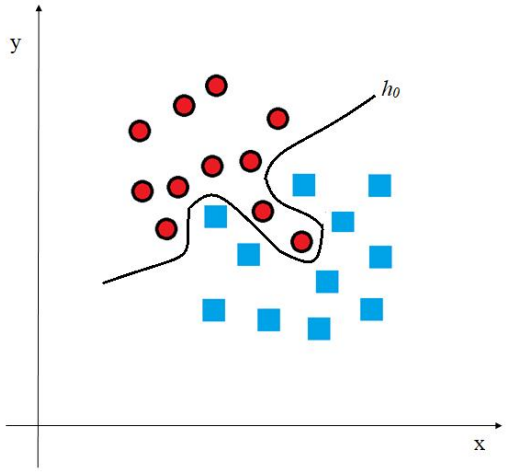

Figura 4.9 - Superfície de separação entre classes em um exemplo não linearmente separável

Para os casos nos quais a solução linear não consegue separar adequadamente as classes, utiliza-se uma classificação não linear dentro de um espaço de dimensões superiores à *n*. Na Figura 4.9 é representado o problema em questão no qual h_0 separa as duas classes existentes.

Funções de Kernel

Para a solução não linear, como indicado anteriormente, é utilizada uma função de *kernel* dada por

$$K(x_i, x'_j) = \phi(x_i)^T \cdot \phi(x'_j),$$

(4.2)

onde ϕ é uma função que transforma um vetor *x* de um espaço *n*-dimensional em um vetor equivalente em espaços de dimensões mais altas (até infinitas).

Essas funções mapeiam os vetores de entrada (exemplos) do treinamento em um espaço dimensional mais alto no qual será encontrada uma solução baseada em um hiperplano (neste novo espaço dimensional).

São três as funções de *kernel* (*K*) mais utilizadas:

- Linear:

  É a função de *kernel* mais simples e é dada por

  $$K(x_i, x'_j) = {x_i}^T \cdot x'_j + b,$$

  (4.3)

  onde *b* é uma constante.

- Polinomial:

  É uma função adequada para problemas nos quais todos os exemplos de treinamento são normalizados e é dada por

  $$K(x_i, x'_j) = (\gamma \cdot {x_i}^T \cdot x'_j + r)^d,$$

  (4.4)

  onde γ, r e d são parâmetros do *kernel* e $\gamma > 0$.

- Função de base radial (do inglês, *Radial Basis Function* - RBF):

  É a principal função de *kernel* pois permite o mapeamento em espaços dimensionais mais altos (melhor que a linear), utiliza uma quantidade de hiper

parâmetros menor que a polinomial e apresenta menores dificuldades numéricas ([BHAVSAR, 2012]). Essa função é dada por

$$K(x_i, x'_j) = e^{(-\gamma \cdot \|x_i - x'_j\|^2)},$$

(4.5)

onde $\gamma > 0$.

4.3.3- *Boosting*

A técnica de *boosting* (o termo pode ser traduzido como melhoramento ou aumento, entretanto essas traduções não são costumeiras na literatura sobre o tema) tem o propósito de transformar classificadores fracos (cujos desempenhos nos processos de classificação são muito ruins) em classificadores fortes (que apresentam resultados quase perfeitos) [ZHOU, 2012].

Essa técnica representa um tipo de método de agrupamento em conjuntos (do inglês, *ensembles*) no qual são feitos treinamentos de múltiplos classificadores e, após isso, os mesmos são combinados gerando *ensembles* que apresentam melhores resultados, considerando as métricas de análise (taxa de erro, acurácia e precisão, por exemplo), que os classificadores aplicados de forma isolada.

O agrupamento de classificadores pode ser feito, basicamente, de forma sequencial ou paralela. O processo de agrupamento realizado de forma paralela (na qual o treinamento dos classificadores envolvidos ocorre paralela ou concomitantemente) também é chamado de ensacamento (do inglês, *bagging*) e não é alvo deste trabalho. Já o processo de criar *ensembles* a partir de treinamentos sequenciais é chamado de *boosting* [SIMEONE, 2017] e é adotado nesta tese para atacar o problema principal estudado.

Um procedimento geral da técnica de *boosting* pode ser representado por meio do Algoritmo 4.1 (adaptado de [ZHOU, 2012]).

Algoritmo 4.1 - Procedimento geral da técnica de *boosting*

ENTRADAS
Distribuição amostral inicial D_1
Classificador base $Class_1$ *% Class*$_1$ *é um classificador fraco*
Número de iterações do processo de boosting N
INÍCIO
Para cada iteração (i) do processo de boosting, fazer:
Treinar o Classificador $Class_i$ *usando a Distribuição amostral* D_i
Avaliar o Erro do Classificador
Ajustar a Distribuição amostral criando D_{i+1}
Definir um novo Classificador $Class_{i+1}$ *mais adequado à Distribuição* D_{i+1}
B =Combinação dos Classificadores {$Class_1$, $Class_2$,..., $Class_n$*}*
FIM
SAÍDA
Classificador B *% B é um classificador forte*

De maneira resumida, o processo de *boosting* ilustrado parte de uma distribuição amostral e um classificador fraco iniciais e, após algumas iterações nas quais são feitos ajustes nas distribuições e nos classificadores com base nos erros de classificação detectados, chega a um classificador forte derivado das combinações dos diversos classificadores que fizeram parte do processo.

O *boosting* tem-se mostrado uma técnica promissora e apresentado bons resultados para problemas de classificação que envolvam o reconhecimento de padrões [KOMORI, 2011]. Por esse motivo, foram escolhidos, para o presente trabalho, dois algoritmos dessa categoria: o *AdaBoost* e o *Robust Boost*.

AdaBoost

O *Boosting* Adaptativo (do inglês, *Adaptative*), ou somente *AdaBoost*, é um *ensemble* que representa um dos principais resultados da aplicação da técnica de *boosting* e tem como principal objetivo, minimizar, para uma classificação binária, a função de perda exponencial [ZHOU, 2012] dada por

$$lo_{exp}(Class|D) = E_{x\sim D}\left[e^{-f(x)Class(x)}\right],$$

(4.6)

onde l_{exp} é a função de perda exponencial, *Class* é o classificador a ser treinado considerando a distribuição amostral *D*, *x* é um exemplo submetido ao processo de treinamento e *f* é a função que retorna o rótulo de *x*.

O algoritmo anterior pode ser adaptado de tal maneira que represente o *AdaBoost*. Neste caso, os passos de "Avaliar o erro", "Ajustar a distribuição" e "Definir novo classificador" devem considerar o problema de minimização apontado anteriormente. Assim, na avaliação do erro, deve-se considerar uma condição de parada que interrompa as iterações caso seja avaliado um erro muito grande (> 0,5), o que torna o algoritmo bastante sensível a ruídos e exemplos discrepantes ou "fora da curva" (do inglês, *outlier*). Além disso, na definição do novo classificador (para a próxima iteração), devem ser considerados os pesos calculados a partir do erro avaliado. Esses pesos, empregados adequadamente na definição do próximo classificador, fazem com que o comportamento desse classificador tenda a ser melhor do que o do classificador anterior no contexto das instancias que foram mal classificadas na iteração ora concluída.

Robust Boost

Considerando o problema da suscetibilidade do *AdaBoost* aos ruídos e *outliers* eventualmente existentes em um conjunto de exemplos de treinamento, optou-se, neste trabalho, por usar, também, um algoritmo de *boosting* mais tolerante a esse tipo de ocorrência. O classificador escolhido, então, foi o *Boosting* Robusto (do inglês, *Robust*) ou simplesmente *Robust Boost*.

Esse algoritmo busca melhorar a tolerância aos ruídos (observada no caso do *AdaBoost*) aumentando a margem normalizada da classificação e, ao invés de minimizar a função de perda exponencial, tenta minimizar a função de erro *E* dada por

$$E_{x\sim D} = \left[I\left(\frac{\sum_{t=1}^{T}\alpha_t f(x)Class_t(x)}{\sum_{t=1}^{T}\alpha_t}\right) \leq \theta\right],$$

(4.7)

onde θ é a margem a ser atingida, *Class* é o classificador a ser treinado e α é o peso dado ao classificador que está sendo treinado.

4.3.4- Procura pelo vizinho mais próximo

A procura pelo vizinho mais próximo (do inglês, *Nearest Neighbor Search* - NNS) ou, em contextos mais genéricos, pelo vizinho de ordem de proximidade *k* (do inglês, *k-Nearest Neighbor* - kNN) fundamenta-se na ideia de que objetos que são semelhantes no espaço de entrada de um processo de classificação também são similares no espaço de saída [ZHOU, 2012].

NNS é um exemplo de classificador que, diferentemente dos demais vistos até aqui, utiliza regras de aprendizado não paramétricas [SIMEONE, 2017], ou seja, que não se baseiam em um modelo paramétrico de relacionamento entre entradas e saídas. Ao invés disso, esse classificador considera que exemplos de treinamento que possuem valores próximos tendem a ter rótulos similares.

Sob uma ótica geométrica, a NNS tem o objetivo de encontrar, em um espaço *n*-dimensional, um ponto, dentre vários pertencentes ao mesmo conjunto de pontos, que seja mais próximo a uma determinada posição de referência.

De maneira formal, a NNS é definida como:

- Dado um conjunto *S* formado por pontos p_i pertencentes a um espaço *n*-dimensional *M* e um ponto de referência *q* também pertencente a *M*, NSS deve encontrar o p_i pertencente a *S* mais próximo a *q*.

Esse processo de classificação não apresenta técnicas de treinamento sofisticadas como os outros algoritmos de ML e muitas vezes se resume a armazenar o conjunto de treinamento (ou um conjunto de valores derivados desse conjunto de treinamento) de tal maneira que, durante a classificação (busca pelo rótulo de uma variável não rotulada), simplesmente realiza uma busca pela menor distância euclidiana entre todos os pontos do conjunto de treinamento e a variável não rotulada.

De maneira análoga aos algoritmos de ML que utilizam a ideia de hiperplano no espaço *n*-dimensional (*M*), a NNS usa o conceito de hiperesfera que representa uma porção de *M* ao redor de um ponto de referência.

Geralmente, quando se diz que o algoritmo de NNS está sendo treinado, estão sendo criadas hiperesferas ao redor de pontos de referência calculados com base nos dados de treinamento. Posteriormente, no momento da classificação, tenta-se verificar se a variável não rotulada está dentro de alguma hiperesfera.

4.4- CONCLUSÃO

O presente capítulo tem o propósito de apresentar fundamentos teóricos sobre a autenticação baseada em eletrocardiografia e sobre o aprendizado de máquina, temas, estes, que são aplicados diretamente na proposta de protocolo que é o foco principal desta tese.

Aqui são apresentados aspectos do funcionamento do coração priorizando os que estão relacionados ao seu funcionamento elétrico e que podem ser registrados por meio de eletrocardiogramas. Além disso, a biometria baseada em eletrocardiografia é explorada de maneira detalhada, inclusive com a apresentação de vários tipos de sensores que podem ser usados para esse propósito. Por fim, são explorados alguns fundamentos e apresentadas as técnicas de aprendizado de máquina que são usadas na proposta de protocolo que é tema central desta tese.

Como conclusão parcial, é possível afirmar que a eletrocardiografia e o aprendizado de máquina podem ser utilizados em conjunto para prover autenticação a usuários de sistemas computacionais que permitam o uso da biometria.

5- PROTOCOLOS DE AUTENTICAÇÃO - TRABALHOS RELACIONADOS

Nesse capítulo são apresentados alguns trabalhos que exploram problemas semelhantes ao que é foco nesta tese. Há trabalhos mais gerais sobre esquemas, protocolos ou modelos de autenticação para ambientes de CC e MCC, e trabalhos mais específicos que empregam, inclusive, a própria eletrocardiografia para a autenticação.

5.1- PROPOSTAS DE ESQUEMAS, PROTOCOLOS OU MODELOS DE AUTENTICAÇÃO PARA AMBIENTES DE CC E MCC

5.1.1- Esquema de autenticação eficiente para serviços de MCC

Em [HE, 2018] é proposto um esquema de autenticação ciente de privacidade (do inglês, *privacy-aware authentication* - PAA) para resolver o problema de identificação em ambientes de MCC.

Uma característica interessante dos sistemas de PAA é o fato das identidades dos participantes poderem ser protegidas, viabilizando a garantia da privacidade desses dados o que, para certas situações, mostra-se um aspecto importante no contexto da MCC.

Um esquema proposto anteriormente [TSAI, 2015] é analisado e são indicadas questões que demonstram que o mesmo, além de ser inseguro contra o ataque de "personificação perante provedor de serviço", não viabiliza a garantia da anonimidade dos usuários.

Os autores, então, propõem um novo esquema que resolve os problemas observados no modelo anterior, fundamentando-se na utilização de *Smart Cards*, em assinaturas baseadas em identidade e no uso de emparelhamento bilinear.

Cabe salientar que, para o uso dos *Smart Cards*, é considerada a utilização de senha e de *fingerprint* a fim de proteger os parâmetros armazenados no cartão.

Para o novo esquema, é proposto o modelo de rede da Figura 5.1 no qual estão presentes as entidades "Usuário", "CSP" e "SCG" (gerador de cartão esperto, do inglês,

Smart Card Generator) e no qual fica visível que os usuários e os CSP realizarão autenticação mútua.

O faseamento do esquema proposto está baseado em três etapas:

1. Fase do configuração (do inglês, *setup*) do sistema - o SCG é configurado escolhendo parâmetros que serão utilizados nas próximas fases;
2. Fase de registro - os usuários e os CSP realizam seu registro perante o SCG e obtêm suas chaves privadas que serão utilizadas na próxima fase;
3. Fase de autenticação - os usuários e CSP autenticam-se mutuamente e geram uma chave de sessão que será utilizada daí para frente para garantir a segurança de suas trocas de mensagens.

Após a apresentação do novo esquema, os autores apresentam o modelo de segurança considerado e fazem uma análise do atendimento aos seguintes requisitos de segurança: autenticação mútua, anonimidade do usuário, não rastreabilidade, estabelecimento de chave, segurança da chave de sessão conhecida, sigilo perfeito à frente, inexistência de tabela de verificação, não sincronização de relógio. Além disso, mostram que o esquema é resistente aos ataques: ataque de dentro, ataque do cartão roubado, ataque de repetição, ataque da personificação do usuário, ataque de *spoofing* do CSP, ataque da tabela de verificação roubada, ataque do homem no meio.

Por fim, é realizada uma análise de desempenho focalizando os custos de computação e de comunicação.

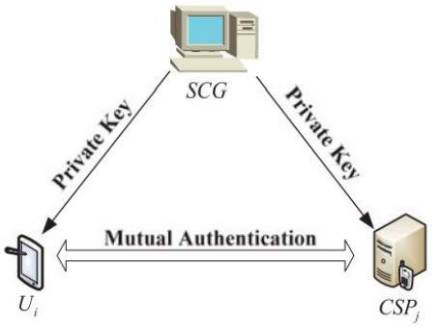

Figura 5.1- Modelo de rede
[HE, 2018]

O esquema proposto em [HE, 2018], apesar de considerar diversos aspectos importantes para evitar os ataques indicados, inclusive com a utilização de biometria baseada em *fingerprint*, não garante que o usuário permaneça autenticado de forma contínua perante o CSP. Caso, por exemplo, um usuário realize a sua autenticação inicial por meio de um computador e saia de perto deste equipamento por alguns instantes, um intruso terá a capacidade de obter acesso aos serviços que seriam direcionados para o usuário autenticado.

Nesse contexto, a autenticação inicial por meio de *fingerprint* não foi suficiente para manter a segurança de maneira contínua, o que é uma das importantes características do protocolo proposto nesta tese.

5.1.2- Esquema melhorado para autenticação eficiente de serviços de MCC distribuídos

Em [XIONG, 2017] é proposto um esquema de autenticação de assinatura única (do inglês, *Single Sign On* - SSO) ciente de privacidade melhorado direcionado aos processos de identificação em serviços distribuídos de MCC.

Esse esquema viabiliza autenticação mútua sem auxílio de um centro de registro online, é resistente a vários ataques conhecidos ("ataque da senha errada", "ataque de personificação"), viabiliza anonimidade dos usuários e sigilo perfeito à frente, além de apresentar vantagens em termos de desempenho (custo computacional e de comunicação) quando comparado a outros esquemas similares.

O esquema é baseado na utilização de emparelhamento bilinear, considera quatro entidades (o usuário, seu equipamento móvel - MD, o centro de registro - RC - que representa uma "terceira parte confiável" - TTP, e o provedor de serviços - SP) e é descrito em cinco fases:

1. Fase de inicialização - o centro de registro (do inglês, *Registration Center* - RC) utiliza emparelhamento bilinear para criar uma chave privada e seus parâmetros vinculados para a troca de mensagens que ocorrerá posteriormente com os usuários;
2. Fase de registro do usuário - o usuário se registra perante o RC enviando e recebendo parâmetros;

3. Fase de registro do provedor de serviços - o provedor de serviços se registra perante o RC de forma análoga à fase anterior;
4. Fase de autenticação - ocorre a autenticação mútua do usuário e do provedor de serviços utilizando os parâmetros obtidos e calculados a partir das três fases anteriores e, ao término, é gerada uma chave de sessão que garantirá segurança às trocas de mensagens que ocorrerão entre o usuário e o provedor;
5. Fase de alteração de senha - ocorre quando o usuário quer alterar sua senha e acontece no próprio equipamento do usuário (as outras entidades não são envolvidas).

É interessante observar que, para a proteção dos parâmetros secretos utilizados pelos usuários durante as fases descritas, são utilizadas apenas senhas criadas pelos próprio usuários (não se utiliza biometria nem dados armazenados em um *Smart Card*).

Finda a descrição, são realizadas análises de segurança e de desempenho. É definido um modelo de segurança a partir do qual é realizada uma demonstração formal da segurança existente no emprego do esquema. Além disso, indica-se que a proposição é resistente aos ataques "de dentro", "verificador roubado", "de repetição", "da personificação do usuário", "do homem no meio", "da senha de login/alteração errada", é baseada em segurança de dois fatores, provê autenticação mútua, anonimidade, não rastreabilidade e sigilo perfeito à frente.

Ao final são realizadas comparações de desempenho com outros esquemas propostos na literatura existente.

A observação feita em relação ao trabalho anterior também cabe para esse trabalho, pois o mesmo não apresenta formas de viabilizar uma autenticação contínua como a proposta desta tese. O fato de se utilizar uma autenticação bimodal é positivo, entretanto a não utilização de características biométricas permite que um atacante, de posse do *smart card* e da senha de um usuário lícito, acesse os serviços aos quais não deveria ter acesso.

5.1.3- Serviço de autenticação para CC baseado em padrões de digitação

Em [ABO-ALIAN, 2016] é proposto um novo sistema de autenticação baseado na dinâmica de digitação (do inglês, *keystroke*) em teclados convencionais e teclados de telas

sensíveis ao toque (do inglês, *touch-screen*), o que é um exemplo da aplicação de biometria comportamental para aumentar a força da autenticação de usuários de serviços de nuvem computacional.

A arquitetura do sistema pode ser compreendida a partir da análise da Figura 5.2 na qual são indicadas as duas fases do processo:

6. Fase de registro do usuário - na qual são realizadas a aquisição de dados por meio do teclado (convencional ou *touch-screen*), a extração das características de *keystroke*, um agrupamento (do inglês, *clustering*) para facilitar a futura busca na base de dados de autenticação, e o armazenamento nessa base;
7. Fase de *login* do usuário - onde se utilizam credenciais convencionais (nome de usuário e senha) para uma primeira verificação do usuário e, caso a verificação seja positiva, procede-se à extração das características biométricas para a comparação com o conteúdo que se encontra armazenado na base de autenticação.

No restante do trabalho, é feita uma análise detalhada do processo de obtenção e análise das características biométricas que serão utilizadas no processo de autenticação.

Não são realizadas análises de segurança e de desempenho.

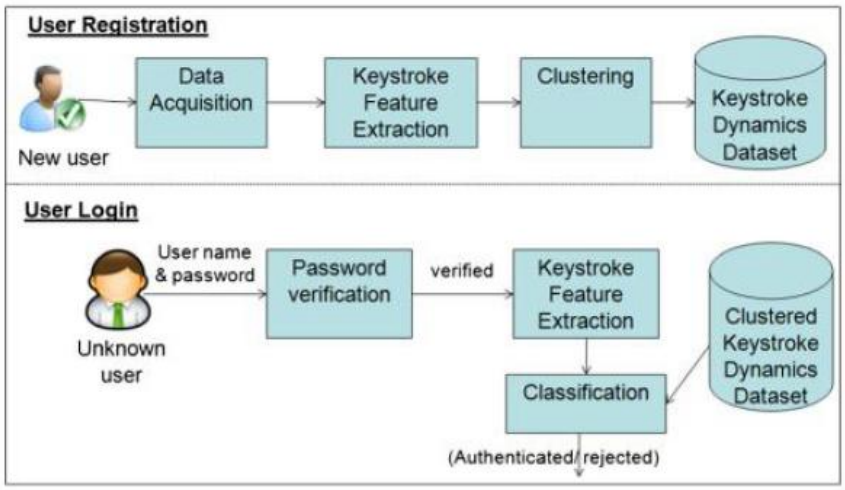

Figura 5.2 - Arquitetura do sistema de autenticação baseado em *keystrokes* [ABO-ALIAN, 2016]

Um comentário que pode ser feito em relação à arquitetura proposta é o fato dela não fazer uso de "algo que o usuário possui" (um *smart card* ou um *token*, por exemplo), que é uma das categorias cujo uso deve ser foco de sistemas de autenticação fortes. Além disso, a técnica de biometria comportamental escolhida, apesar de poder viabilizar uma autenticação contínua do usuário, apresenta problemas quando aplicada à autenticação perante provedores de conteúdo que não demandam tantas ações por parte do usuário (provedores de vídeos e de músicas, por exemplo) [LEE, 2018].

5.1.4- Sistema eficiente de autenticação para MCC baseado em vários fatores

Em [JEONG, 2014] é proposto um novo sistema de autenticação baseado em vários fatores (usuário e senha, reconhecimento de voz, reconhecimento de face, número de identidade internacional do equipamento móvel - IMEI e número de identidade internacional do assinante móvel - IMSI) para prover segurança de forma eficiente em ambientes de computação móvel em nuvem.

Um dos aspectos interessantes do protocolo é o fato das análises de cada um dos fatores utilizados para a autenticação serem realizadas em máquinas virtuais diferentes disponibilizadas pelo ambiente de nuvem computacional, o que viabiliza a paralelização do processo e a conseguinte melhoria de sua eficiência.

Antes de descrever o sistema propriamente dito, o trabalho indica uma série de considerações de segurança da informação que precisam ser observadas para um contexto de MCC e sugere a utilização de nuvens computacionais híbridas (parte públicas, parte privadas) com o objetivo de fornecer segurança adequada aos vários tipos de dados existentes (ostensivos e sigilosos). Além disso, sinaliza aspectos de segurança que devem ser observados no contexto dos equipamentos móveis que são utilizados no ambiente de MCC.

Como critérios resumidos para validação de um processo de autenticação, o trabalho sinaliza que as entradas de usuários devem ser fáceis de serem lembradas e disponibilizadas (digitadas, etc.), e devem ser difíceis de serem perdidas, além disso, as solicitações de autenticação devem ser processadas rapidamente.

A descrição da arquitetura do sistema pode ser observada na Figura 5.3 na qual são percebidas as seguintes características:

- Os cinco fatores de autenticação (usuário/senha, voz, face, IMEI e IMSI) capturados pelo equipamento móvel do usuário são submetidos a uma função *hash* dentro do equipamento;
- A comunicação entre o equipamento móvel e o ambiente de nuvem é protegida por protocolos de segurança convencionais utilizados em redes sem fio (WPA2, etc.) e em transações via Internet (SSL/TLS, etc.);
- Os dados de cada um dos fatores de autenticação considerados são armazenados em áreas separadas da base de autenticação;
- Cada fator é analisado em um máquina virtual específica para fins de aumento da eficiência.

A operação do sistema dá-se em duas fases das quais participam as entidades "equipamento do usuário", "servidor de gerenciamento", "área de armazenamento" e "área de processamento":

1. Registro - os cinco fatores de autenticação são extraídos (pode ser utilizado o próprio equipamento do usuário ou outros que viabilizem a extração adequada das características necessárias) e armazenados na área de armazenamento (para que possíveis redundâncias sejam bloqueadas, os dados a serem inseridos deverão ser comparados aos já armazenados);
2. Autenticação - os cinco fatores são capturados a partir do equipamento do usuário (utilizando os sensores e demais componentes do próprio aparelho) e são enviados para as cinco máquinas virtuais responsáveis pelos processos de comparação a serem feitos entre dados capturados e dados armazenados; caso todos os processos sinalizem êxito em suas comparações, o processo de autenticação será concluído com êxito, do contrário, havendo qualquer comparação errada, a autenticação irá falhar.

Após a descrição do novo sistema, os autores apresentam análises abreviadas de segurança, desempenho e expansibilidade utilizando comparações com outros trabalhos similares.

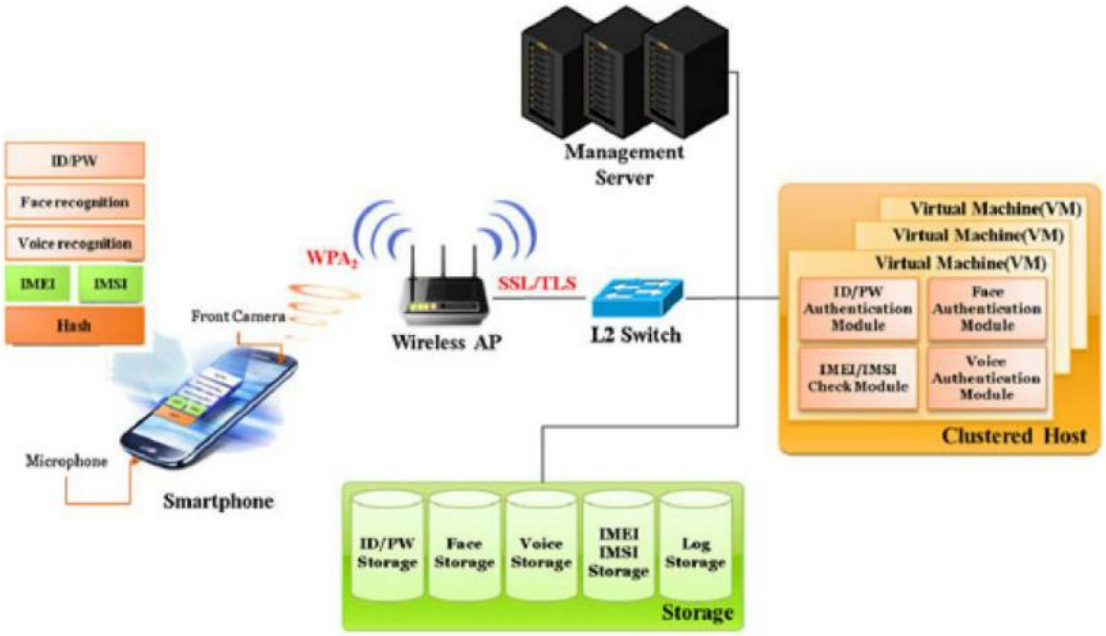

Figura 5.3 - Arquitetura do sistema
[JEONG, 2014]

O sistema proposto possui algumas partes que podem ser consideradas desnecessárias. A utilização de IMEI e IMSI como fatores de autenticação pode ser observada como algo redundante (e por esse motivo, que atenta contra aspectos de eficiência), haja vista que ambos representam um fator de autenticação "que se tem". Além disso, o IMEI impede que o usuário utilize outros equipamentos para acesso aos serviços providos pelo CSP. Caso a opção fosse pela utilização apenas do IMEI (como o protocolo proposto na presente tese), o usuário poderia utilizar qualquer equipamento, pois o IMEI encontra-se no *chip* que pode ser passado de um UE para outro.

Sobre os tipos de autenticação biométrica utilizados (reconhecimento de voz e de face), esses apresentam desvantagens (como indicado no item 4.2.1-"Aspectos gerais") em relação à eletrocardiografia, que é a opção adotada nesta tese.

5.1.5- Comparação entre os trabalhos apresentados

A presente comparação é feita sobre os trabalhos descritos no item 5.1-"Propostas de esquemas, protocolos ou modelos de autenticação para ambientes de CC e MCC" e o proposto nesta tese.

A Tabela 5.1 apresenta uma comparação entre os diversos esquemas, protocolos ou modelos apresentados e o protocolo proposto nesta tese.

Tabela 5.1 - Comparação entre os trabalhos gerais analisados

Características/ Trabalhos	**HE, 2018**	**XIONG, 2017**	**ABO-ALIAN, 2016**	**JEONG, 2014**	**Proposta desta tese**
Tipo de nuvem	MCC	MCC	CC	MCC	MCC
autenticação mútua	Sim	Sim	Não	Não	Sim
autenticação contínua	Não	Não	Sim	Sim	Sim
smart card/SIM *card*	Sim	Não	Não	Sim	Sim
necessita de TTP	Não	Sim	Sim	Não	Não
Fatores de autenticação	Senha, impressão digital	Senha	Senha, análise de digitação	Senha, reconhecimento de voz, reconhecimento facial, IMSI, IMEI	Senha, ECG, IMSI

5.2- PROPOSTAS DE UTILIZAÇÃO DE ELETROCARDIOGRAFIA PARA IDENTIFICAÇÃO E AUTENTICAÇÃO

5.2.1- Reconhecimento biométrico de ECG utilizando uma abordagem baseada em SVM

Em [REZGUI, 2016] é proposta uma abordagem para a identificação de pessoas com base em seus eletrocardiogramas. Nesse trabalho são escolhidas 21 características (do inglês, *features*) derivadas de pontos fiduciais (distâncias em tempo e em tensão entre os pontos) e 10 descritores morfológicos obtidos da análise de cada ciclo cardíaco.

O trabalho faz uma crítica a outros trabalhos relacionados indicando que os mesmos são fortemente baseados em abordagens que consideram somente as localizações dos pontos fiduciais, o que gera atributos que representam apenas algumas especificidades dos eletrocardiogramas. Por esse motivo, os autores consideram descritores morfológicos que agregam parâmetros de análise ao processo como um todo.

São utilizados ECG de duas bases de dados públicas encontradas em [MIT-BIH, 2000a] e [MIT-BIH, 2000b] que contêm, somadas, 66 registros.

O sistema de identificação biométrica proposto é desencadeado em três passos: o pré-processamento, a extração das *features* e a classificação.

1. O primeiro passo tem o propósito de filtrar o sinal de ECG de tal forma que sejam retirados, das análises, os ruídos presentes no eletrocardiograma. Baseado em resultados empíricos, os autores optam por utilizar frequências de corte para o filtro passa banda (do inglês, *band-pass*) de estão entre 2 e 40 Hz.
2. Para a extração dos pontos fiduciais, o trabalho indica a utilização de um detector automático de "onda QRS" encontrado na literatura que, de acordo com a análise dos próprios autores, não apresenta boa acurácia. Esse detector permite, mesmo assim, a captura dos elementos necessários para o cálculo das *features* (derivadas dos pontos fiduciais) e dos descritores morfológicos que são adotados conforme a Figura 5.4.
3. A classificação, último passo do processo, está baseada em 44 máquinas de vetores de suporte (SVM) que são empregadas em classificações binárias, uma para cada sujeito. São testados dois diferentes *kernels*, o RBF e o polinomial, e uma das conclusões aponta para melhores resultados do *kernel* RBF.

De acordo com as tabelas indicadas no artigo, o melhor resultado obtido pelo sistema de identificação foi uma taxa de reconhecimento de 99,38% a partir da utilização dos 31 parâmetros sobre os registros da base de [MIT-BIH, 2000b].

Apesar do texto inicial indicar a utilização de SVM para classificar sujeitos, a divisão que os autores apresentam referente às classes de patologias existentes nas bases gera dúvidas em relação ao real propósito da classificação: identificar indivíduos ou classes de problemas cardíacos. Outro aspecto controverso é o fato dos autores afirmarem que existem registros de 66 indivíduos (considerando as duas bases de ECG utilizadas) e só existem 44 classificadores binários.

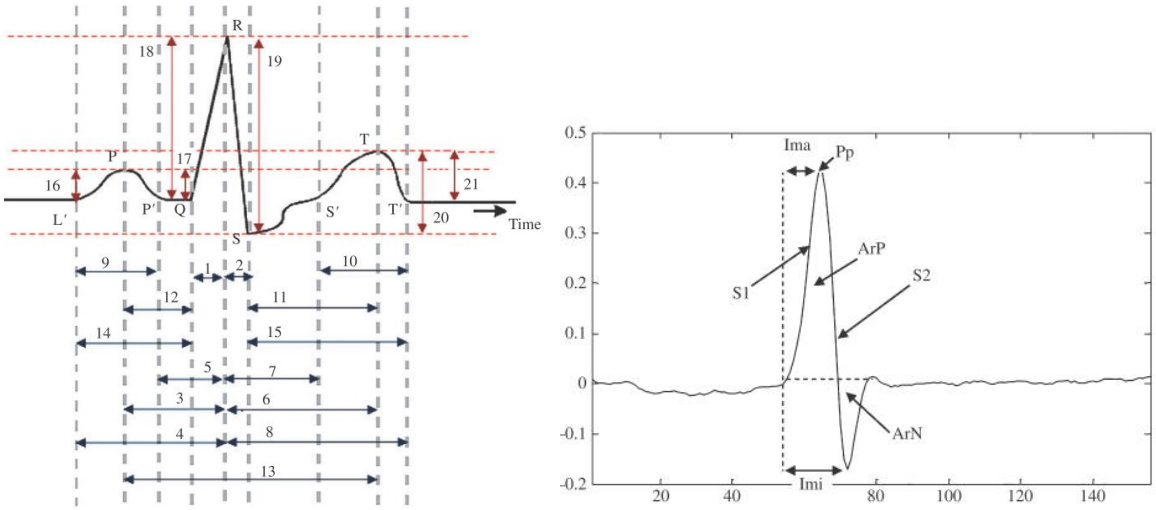

Figura 5.4 - *Features* extraídas dos pontos fiduciais (à esquerda) e descritores morfológicos (à direita)
[REZGUI, 2016]

Um problema observado neste artigo diz respeito ao detector automático de QRS que, de acordo com as avaliações dos próprios autores, é pouco acurado. Considerando que todas as classificações posteriores são realizadas com base naquilo que é detectado pelo primeiro algoritmo (detector automático de QRS), caso haja imprecisões neste processo, todo o restante da análise poderá estar comprometido.

Outro aspecto que pode ser comentado é a utilização de "*features* derivadas de outras *features*", o que indica o emprego redundante de informações e provável aumento desnecessário das análises feitas. A Figura 5.4 (esquerda) mostra, por exemplo, que as *features* 9 e 5, somadas, são iguais à *feature* 4; 1 e 12, somadas, são iguais à 3; 2 e 11, somadas, são iguais à 6. Da mesma forma, alguns descritores morfológicos (à direita, na figura) derivam das próprias *features*. É o caso, por exemplo, de descritores que representam áreas sob as curvas. Essas áreas são calculadas utilizando-se as distâncias representadas por algumas *features*.

5.2.2- Autenticação para equipamentos móveis baseada em ECG

Em [FALCONI, 2016] é proposto um novo algoritmo de autenticação biométrica baseado em eletrocardiografia e voltado para a computação móvel (os autores afirmam que esse é o primeiro algoritmo desse tipo publicado na literatura).

A abordagem considerada para a autenticação baseada em ECG focaliza, nesse artigo, a utilização de 8 *features* derivadas de pontos fiduciais (distâncias em tempo e em potencial elétrico entre os pontos) que não apresentam redundâncias.

São utilizados 73 registros de usuários de 4 bases de ECG públicas ([MIT-BIH, 2000a], [MIT-BIH, 2000b], [TADDEI, 1992], [LAGUNA, 1997]) e 10 registros capturados por meio de um sensor com 2 eletrodos (cada eletrodo fica em contato com um dedo de uma mão diferente) utilizado no laboratório dos próprios autores. Esse sensor, de acordo com o artigo, pode ser colocado na capa de um *smartphone*.

O algoritmo é dividido em seis fases: a detecção dos pontos fiduciais, o alinhamento, a normalização, a extração das *features*, a geração de padrões (do inglês, *templates*) a serem armazenados e a serem autenticados, e a comparação:

1. Na primeira fase são detectados os pontos fiduciais (conforme Figura 5.5) que serão utilizados nos cálculos das *features* escolhidas;
2. A segunda fase diz respeito ao alinhamento dos diversos ciclos observados em torno dos seus pontos R, ou seja, todos os pontos R são colocados sobre a mesma posição e os outros pontos, de cada ciclo, são alinhados de tal maneira que haja apenas um ponto de referência a partir do qual todas as medições sejam feitas;
3. Em seguida, é feita uma normalização linear na qual as distâncias entre cada ponto do ECG e o ponto R são divididas pelo comprimento total de um batimento cardíaco, ou seja, a distância entre pontos R consecutivos (essa normalização somente é aplicada às distâncias medidas em tempo). Essa normalização tem o propósito de suprimir os efeitos das variações das frequências cardíacas;
4. Na quarta fase é feita a extração das *features* desejadas por meio de subtrações entre as posições dos diversos pontos fiduciais, em comprimento e em amplitude (ou seja, em tempo e em diferença de potencial elétrico), conforme a Figura 5.5;
5. Extraídas as *features*, parte-se para a geração dos padrões (para armazenamento ou para autenticação) que serão comparados um ao outro (o armazenado, gerado em momentos anteriores, é comparado ao padrão a ser autenticado, que foi gerado posteriormente).
6. Gerados os padrões, é feita a comparação na qual é verificado se o padrão a ser autenticado pertence à mesma classe (ao mesmo usuário) do padrão armazenado ou

não. Essa comparação é feita confrontando (com base em tolerâncias específicas) cada uma das *features* do padrão gerado para a autenticação com as *features* correspondentes do padrão armazenado.

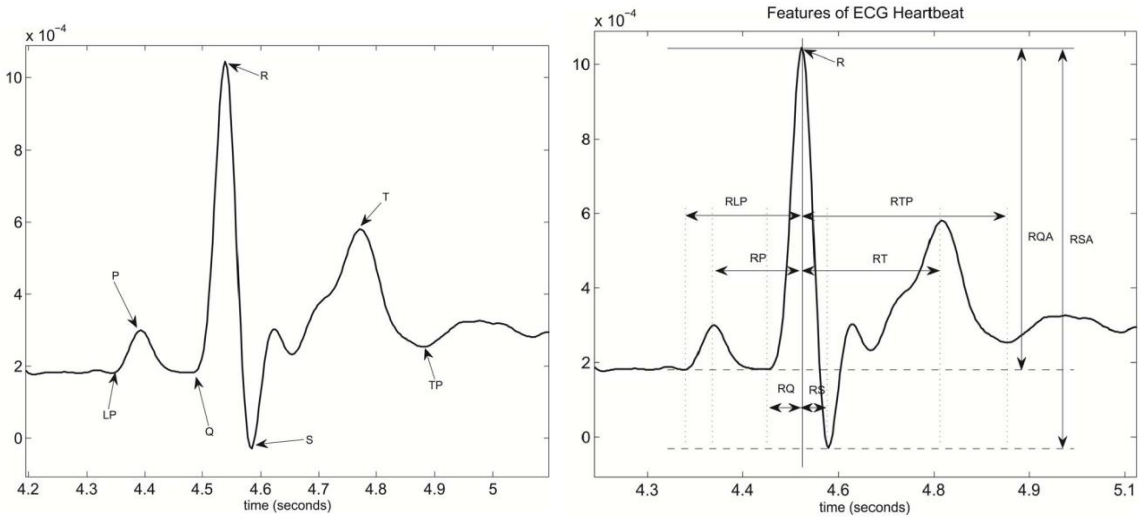

Figura 5.5 - Pontos fiduciais detectados pelo algoritmo (à esquerda) e *features* calculadas (à direita)
[FALCONI, 2016]

De acordo com os dados indicados pelos autores, os melhores resultados obtidos por meio do algoritmo foram: da aplicação do algoritmo sobre dados das bases públicas - TAR igual a 84,93% e FAR igual a 1,29%; da aplicação do algoritmo sobre dados obtidos em laboratório - TAR igual a 81,82% e FAR igual a 1,41%.

Algumas questões merecem ser comentadas em relação ao algoritmo analisado.

- Os autores não deixam claro como realizam a detecção dos pontos fiduciais (não é possível saber se é uma detecção automática ou se é realizada por meio de inspeção visual ou algo do tipo).
- A fase de alinhamento é um aspecto avaliado como desnecessário no algoritmo proposto, pois ela é realizada, de qualquer forma, implicitamente no momento em que as *features* forem calculadas por meio das subtrações indicadas na quarta fase do processo.
- A normalização proposta pelos autores com base na distância entre pontos R consecutivos (comprimento total de um batimento), apesar de sustentada por referências apontadas no próprio trabalho, é algo que pode ser contestado

observando [MASON, 2016], pois as variações nas distâncias temporais entre os pontos fiduciais (QRS, por exemplo), apesar de ocorrerem no mesmo sentido da variação da distância R-R (quando esta aumenta, as outras também aumentam), não são proporcionais à variação dessa distância. Para essa normalização, é o caso ser escolhida outra distância que apresente uma proporcionalidade mais adequada para esse tipo de uso.

- A comparação que os autores utilizam na sexta fase do algoritmo pode, na verdade, ser interpretada como a aplicação de um classificador do tipo NNS.

5.2.3- Utilização de fluxos de eletrocardiogramas em tempo real para a autenticação contínua

Em [CAMARA, 2017] é proposto um sistema de autenticação contínua baseado em ECG e voltado para aplicações em tempo-real. Os autores indicam que tem sido observada uma evolução muito grande na área de sistemas biométricos e que a variedade de biossinais (do inglês, *biosignals*) à disposição para esses sistemas está cada vez maior. Além disso, também é proposta a utilização, para viabilizar bons resultados oriundos dos sistemas biométricos, de técnicas de ML e de mineração de fluxos de dados (inglês, *Data Stream Mining* - DSM).

O sistema funciona a partir da estrutura geral apresentada na Figura 5.6. e segue os seguintes passos:

1. Fase de configuração:
    a. Capturar registros de ECG;
    b. Realizar pré-processamento e extração das *features*;
    c. Criar um modelo de referência para cada usuário.
2. Fase de operação:
    a. Capturar registros de ECG;
    b. Realizar pré-processamento e extração das *features*;
    c. Descartar eventuais amostras que sejam consideradas atípicas por meio de um módulo de similaridade;
    d. Autenticar as amostras por meio do módulo de aprendizagem;
    e. Atualizar o módulo de aprendizagem.

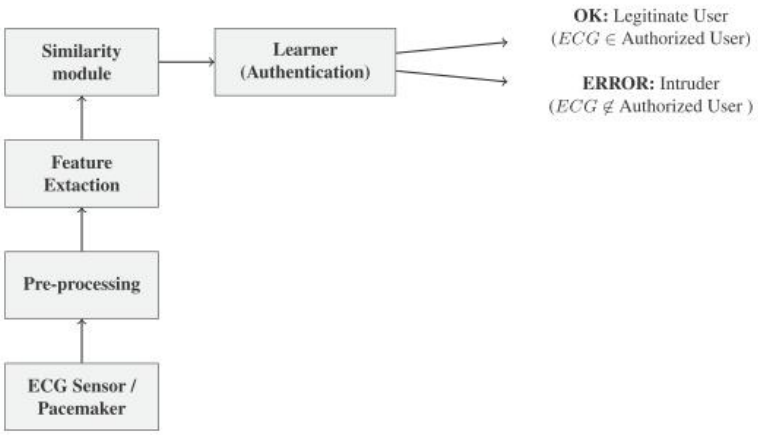

Figura 5.6 - Estrutura geral do sistema de autenticação baseada em ECG [CAMARA, 2017]

Para a fase de registro, os autores indicam a necessidade de realizar uma captura de ECG durante 30 minutos para cada usuário do sistema para, posteriormente, realizar os outros passos (pré-processamento e geração dos modelos de referência).

A abordagem utilizada para a extração das *features*, diferentemente dos outros trabalhos analisados (que utilizam abordagens "fiduciais", ou seja, baseadas nos pontos fiduciais), é considerada "não fiducial" e está fundamentada na obtenção de *features* a partir da aplicação da transformada de "Walsh-Hadamard". Já para o processo de aprendizagem, é utilizada a técnica de aprendizado de máquina kNN.

O treinamento dos classificadores é feito de forma contínua e utiliza um esquema de janelas deslizantes (do inglês, *sliding windows*) que permite uma atualização constante dos parâmetros dos classificadores.

Os experimentos são feitos utilizando os ECG de 10 indivíduos que não apresentam problemas cardíacos e cujos registros encontram-se em uma base de ECG pública ([MIT-BIH, 2000a]). Os resultados finais das análises indicam uma acurácia média de 94,79%.

Algumas observações devem ser feitas em relação ao artigo analisado.

- Os autores afirmam que, neste artigo, é indicada a utilização, pela primeira vez, de análise de fluxos de dados de ECG, não apenas ECG isolados, entretanto, observando os outros artigos analisados, percebe-se que essa ideia já foi aplicada em momentos anteriores.
- A proposição de um módulo de similaridade que tem a finalidade de descartar

amostras que não sigam o modelo de referência do usuário é algo controverso, pois esse módulo pode causar distorções no processo de aprendizagem caso o primeiro modelo de referência não represente fielmente o usuário. De certa forma, esse módulo pode prejudicar o processo de treinamento continuado que é proposto pelos próprios autores, pois cria uma dependência direta com o primeiro modelo de referência adotado, impedindo uma evolução do modelo a partir do descarte de amostras possivelmente representativas.

- Os autores propõem a captura inicial, para cada usuário, de um ECG de 30 minutos de duração. Além de esse método ser, no mínimo, "desconfortável" para os usuários, ele não assegura que, nesses 30 minutos, haverá informação representativa o suficiente para gerar os modelos de referência necessários para o funcionamento do sistema de autenticação.

5.2.4- Autenticação bimodal baseada em ECG e impressão digital

Em [FALCONI, 2018], os mesmos autores de [FALCONI, 2016] propõem uma evolução da proposta anterior acrescentando, ao algoritmo original, outro fator de autenticação baseado em impressões digitais. Além disso, a abordagem empregada em 2016 para a classificação dos dados obtidos dos ECG, que era baseada em NNS, é substituída por uma que utiliza SVM.

A Figura 5.7 indica que, para as duas técnicas biométricas, é feita a captura e a extração das *features* que serão utilizadas na autenticação. Após isso, em ambos os casos, é feita a análise das *features* obtidas no momento da autenticação para verificar se elas pertencem à mesma classe que as *features* armazenadas na base de dados e que representam os padrões dos usuários. Por fim, os resultados das duas análises são utilizados em um algoritmo de tomada de decisão em que, caso pelo menos uma das técnicas indique uma autenticação positiva, o usuário será considerado autêntico.

Para a análise dos ECG, são usadas as mesmas bases de dados públicas de ECG que as utilizadas anteriormente. Esses ECG são divididos em fragmentos de 60 segundos para o treinamento das SVM e de 4 segundos para os testes de classificação. Ao término do processo de avaliação, obtém-se uma TAR de 100% e uma FAR de 7%. Após a aplicação da segunda técnica biométrica, esses valores tornam-se ainda melhores, com a TAR mantendo o valor de 100% e com a FAR diminuindo para 2,96%.

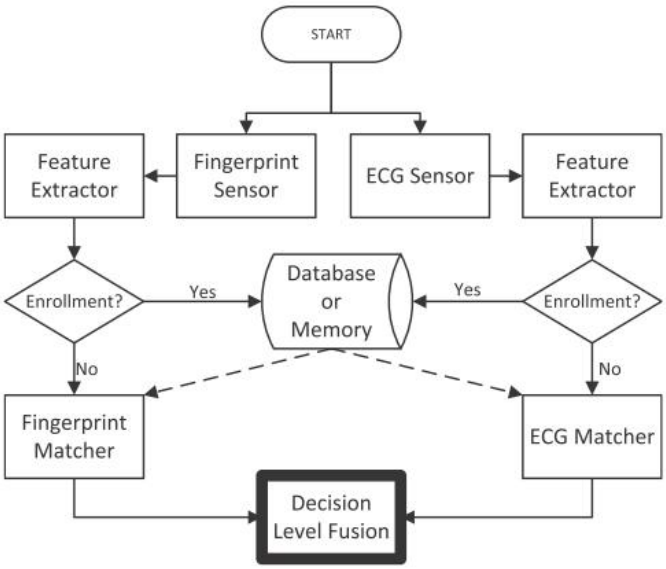

Figura 5.7 - Algoritmo de autenticação bimodal
[FALCONI, 2018]

As mesmas observações (exceto a última) indicadas no trabalho dos autores datado de 2016 podem ser feitas em relação à versão de 2018. A aplicação das SVM, entretanto, mostra uma evolução considerável nos resultados das métricas analisadas.

Além das observações indicadas, merece menção o fato do algoritmo de autenticação bimodal indicado impor uma constante utilização da técnica baseada em impressões digitais (as duas técnicas são executadas de forma paralela). Esse modelo de autenticação, então, não pode ser considerado não intrusivo e contínuo.

5.2.5- Comparação entre os trabalhos apresentados

A presente comparação é feita sobre os processos descritos no item 5.2-"Propostas de utilização de eletrocardiografia para identificação e autenticação" e o proposto nesta tese.

A Tabela 5.2 apresenta uma comparação entre as diversas propostas de utilização de eletrocardiografia apresentadas e a proposta nesta tese.

Tabela 5.2 - Comparação entre os trabalhos analisados

Características/ Trabalhos	**[REZGUI, 2016]**	**[FALCONI, 2016]**	**[CAMARA, 2017]**	**[FALCONI, 2018]**	**Proposta desta tese**
Fatores de autenticação	ECG	ECG	ECG	ECG e impressão digital	ECG, senha e IMSI
Abordagem	Detecção de pontos fiduciais	Detecção de pontos fiduciais	Utilização da transformada de Walsh-Hadamard	Detecção de pontos fiduciais (para o ECG)	Detecção de pontos fiduciais
Tipo de classificador	SVM	NNS	kNN (k=1)	SVM (para o ECG)	SVM, *AdaBoost*, *Robust Boost* e NNS
Atualização dos classificadores	Não aborda	Não aborda	*Sliding windows*	Não aborda	Fase própria de atualização
Número de parâmetros do classificador	31	8	Não informado	8 (ECG)	8
Dados são normalizados?	Assunto não tratado	Normalização com base em R-R	Não se aplica	Normalização com base em R-R (ECG)	Normalização com base em LP-TP
Base de ECG utilizada	[MIT-BIH, 2000a] e [MIT-BIH, 2000b]	[MIT-BIH, 2000a], [MIT-BIH, 2000b], [TADDEI, 1992], [LAGUNA, 1997]	[MIT-BIH, 2000b]	[MIT-BIH, 2000a], [MIT-BIH, 2000b], [TADDEI, 1992], [LAGUNA, 1997]	Base indicada em [SOUZA, 2019]
Número de usuários analisados	66	73	10	73	108
Acurácia max.	99,38%	81,82%	94,79%	100%	94,7%
Precisão max.	-	-	-		99,5%
Sensibilidade Max	-	-	-		97,1%
F1-Score Max	-	-	-		94,9%

É importante frisar que não foi possível realizar uma comparação mais precisa do protocolo proposto nesta tese com as demais propostas constantes nos trabalhos indicados, pois, para essa comparação, todos os processos deveriam ser aplicados sobre a mesma base de ECG, o que se mostrou inviável nas duas situações possíveis:

1. Opção pela base indicada em [SOUZA, 2019] - apesar da proposta desta tese ter considerado essa base, ela não foi utilizada nos outros trabalhos e, além disso, os autores não indicaram, em seus artigos, detalhes suficientes dos processos propostos para que fosse possível reimplementá-los de forma exata (tal qual os autores dizem ter implementado) no escopo deste trabalho, o que não permitiu a obtenção de valores de métricas dos outros processos sobre essa base de ECG;
2. Opção pelas demais bases indicadas na Tabela 4.1 - considerando que essas bases apresentam ECG de muitos usuários que possuem problemas cardíacos (arritmias, dentre outros), não foi possível extrair dados adequados (em quantidade e qualidade) utilizando o processo de detecção de pontos fiduciais implementado para esta tese (será detalhado no próximo capítulo) e, dessa forma, não foi possível aplicar o protocolo proposto utilizando dados extraídos dessas bases, o que também inviabilizou uma comparação mais adequada entre as métricas extraídas da aplicação das propostas.

5.3- CONCLUSÃO

O presente capítulo tem o propósito de apresentar alguns trabalhos atuais encontrados em importantes referências bibliográficas e que estão relacionados às áreas que são exploradas nesta tese.

São apresentados dois tipos de propostas, um que focaliza a autenticação para ambientes de CC e MCC, e outro que aborda a utilização da eletrocardiografia para a autenticação de seres humanos. As principais ideias dos trabalhos são apresentadas para cada um dos tipos e são criadas duas tabelas comparativas dos trabalhos entre si e em relação à proposta feita nesta tese.

Observando as tabelas referentes aos dois tipos indicados e os comentários sobre cada um dos trabalhos, pode-se concluir que a proposta central desta tese apresenta uma série de vantagens em relação às propostas encontradas nos trabalhos relacionados.

6- PROPOSTA DE PROTOCOLO DE AUTENTICAÇÃO MULTIMODAL CONTÍNUA

6.1- CONSIDERAÇÕES INICIAIS

Após explorar as várias bases conceituais necessárias para fundamentar este trabalho, o presente capítulo tem, como principal objetivo, apresentar uma proposta de protocolo para prover autenticação multimodal, não intrusiva, e contínua para usuários de ambientes de computação móvel em nuvem - MCC.

A Tabela 6.1 apresenta as variáveis, constantes e funções utilizadas neste capítulo.

Esse sistema utiliza três fatores para realizar a autenticação dos usuários:

- "algo que a pessoa conhece" - uma senha;
- "algo que a pessoa possui" - um cartão com o Módulo de Identidade do Assinante (do inglês, *Subscriber Identity Module card* - SIM *card*) contendo um IMSI;
- "algo que é intrínseco à pessoa" - seus ciclos cardíacos representados, nesse contexto, por um eletrocardiograma.

A opção por utilização de uma senha é a forma mais simples de representar a primeira categoria de fatores. Como esta técnica será combinada a outros mecanismos, as vulnerabilidades citadas na fundamentação teórica desta tese tendem a ter seus efeitos minimizados. As regras de criação dessa senha também não são alvo de análise aprofundada, sendo adotado um padrão com tamanho mínimo de oito caracteres alfanuméricos (letras maiúsculas e minúsculas, números e símbolos obtidos de um teclado comum) criados a partir da interface do UE.

Em relação à utilização do IMSI, uma alternativa diferente seria a utilização do IMEI, entretanto essa opção vincularia o processo de autenticação a um equipamento específico, o que diminuiria a flexibilidade do sistema restringindo a autenticação do usuário a esse equipamento além de acarretar problemas de segurança vinculados à perda ou roubo do equipamento. Optando-se pelo IMSI, a autenticação do usuário poderá ocorrer a partir de qualquer equipamento que possua as características adequadas e no qual o IMSI esteja presente.

Tabela 6.1 - Variáveis utilizadas neste capítulo

Variável/ constante/ função	**Descrição**
Pwd	Senha do usuário
H( )	Aplicação de uma função de condensação sobre um conteúdo
IdAut	Identificador de autenticação da sessão de comunicação segura entre o CSP e o usuário
$E_0[t]$	Vetor de dados brutos de ECG obtido do HS e formado por valores de amplitudes do ECG que representam potenciais elétricos (em mV)
$E_1[t]$	Vetor de dados filtrados (atenuação de ruído) de ECG
IDX	Vetor que contem os índices de $E_1[t]$ correspondentes aos picos e vales detectados
c	Índice do ciclo cardíaco considerado na análise
LP[c], P[c], Q[c], R[c], S[c], T[c] e TP[c]	Vetores contendo os pontos fiduciais detectados em cada ciclo *c*
RLP[c], RP[c],RQ[c], RS[c], RT[c] e RTP[c]	Vetores contendo as distâncias, em termos de índices, entre os pontos considerados do ciclo *c*
RSA[c] e RQA[c]	Vetores contendo as distâncias, em termos de amplitudes, entre os pontos considerados do ciclo *c*
$f_1[c], f_2[c],.. , f_8[c]$	Vetores de *Features* extraídas para cada ciclo *c*
NC	Número de ciclos considerados para o treinamento do classificador
μ_{fj}	Média da "j-ésima" *feature* considerando todos os ciclos do treinamento
Tol_j	Tolerância referente à "j-ésima" *feature* para uma classificação com NNS
Acertos	Quantidade de acertos na comparação entre *features* e médias armazenadas
AcertosMin	Quantidade mínima de acertos adequada para o processo

Acerca da utilização dos dados de ciclos cardíacos (presentes no ECG) como *features* biométricas da autenticação, essa técnica pode ser considerada não intrusiva e transparente para o usuário (a colocação de um sensor cardíaco de peito como o encontrado em [QARDIO CORE, 2019], por exemplo, pode assemelhar-se à utilização de um simples monitor cardíaco de cinta comumente utilizado por atletas e pode provê os dados necessários para a criação de um ECG representativo para o processo de autenticação). Essa técnica apresenta vantagens de segurança em relação a outras técnicas biométricas como, por exemplo, o reconhecimento de voz, que está sujeito a um ataque baseado na reprodução de uma gravação da voz do usuário; o reconhecimento facial, que está sujeito a

um ataque baseado na apresentação de uma imagem capturada do usuário; e o reconhecimento de impressão digital, que não pode ser considerado transparente para o usuário.

Ainda sobre a utilização da eletrocardiografia, a classificação das *features* biométricas será feita com a utilização de técnicas de aprendizado de máquina (ML).

É importante salientar que, além dos fatores voltados para a autenticação do usuário perante o CSP, esse provedor também é autenticado frente ao usuário por meio de um certificado digital que se encontra no próprio CSP. Dessa forma, o protocolo proposto pode ser considerado de autenticação mútua entre usuário e provedor. Apesar de ser importante, essa característica não será explorada a fundo no presente trabalho pelo fato de não apresentar aspectos inovadores. Nos modelos apresentados à frente, esse aspecto será representado, de forma simplificada, por uma conexão SSL/TLS criada com base no certificado digital do CSP e que protege os fluxos de dados existentes entre o provedor e o equipamento móvel do usuário.

6.2- DEFINIÇÃO DE MODELO DE REDE

A Figura 6.1 representa o modelo de rede no qual o protocolo de autenticação proposto deve funcionar.

As principais entidades integrantes do modelo são:

- Sensor cardíaco (do inglês, *Heart Sensor* - HS) - captura sinais cardíacos do usuário e os fornece para a criação de um ECG representativo do indivíduo;
- Equipamento do usuário (UE) - por meio desse equipamento, o usuário realiza sua autenticação e recebe os serviços disponibilizados pelo provedor de serviços;
- Provedor de serviços de computação em nuvem (CSP) - essa entidade centraliza as solicitações e o fornecimento de serviços da nuvem computacional e, para tanto, necessita verificar a autenticidade dos demandantes (usuários) além de também precisar autenticar-se perante esses (como indicado anteriormente, para esse modelo, o CSP possui um certificado digital vinculado a chaves criptográficas assimétricas que permite a sua autenticação perante os usuários);
- Centro de autenticação contínua (do inglês, *Continuous Authentication Center* -

CAC) - conjunto de entidades (máquinas reais ou virtuais) da nuvem computacional alocadas para as várias tarefas relacionadas aos processos de autenticação;

- Base de dados de autenticação (do inglês, *Authentication Database* - ADB) - local onde são armazenados os dados a serem utilizados nos processos de autenticação.

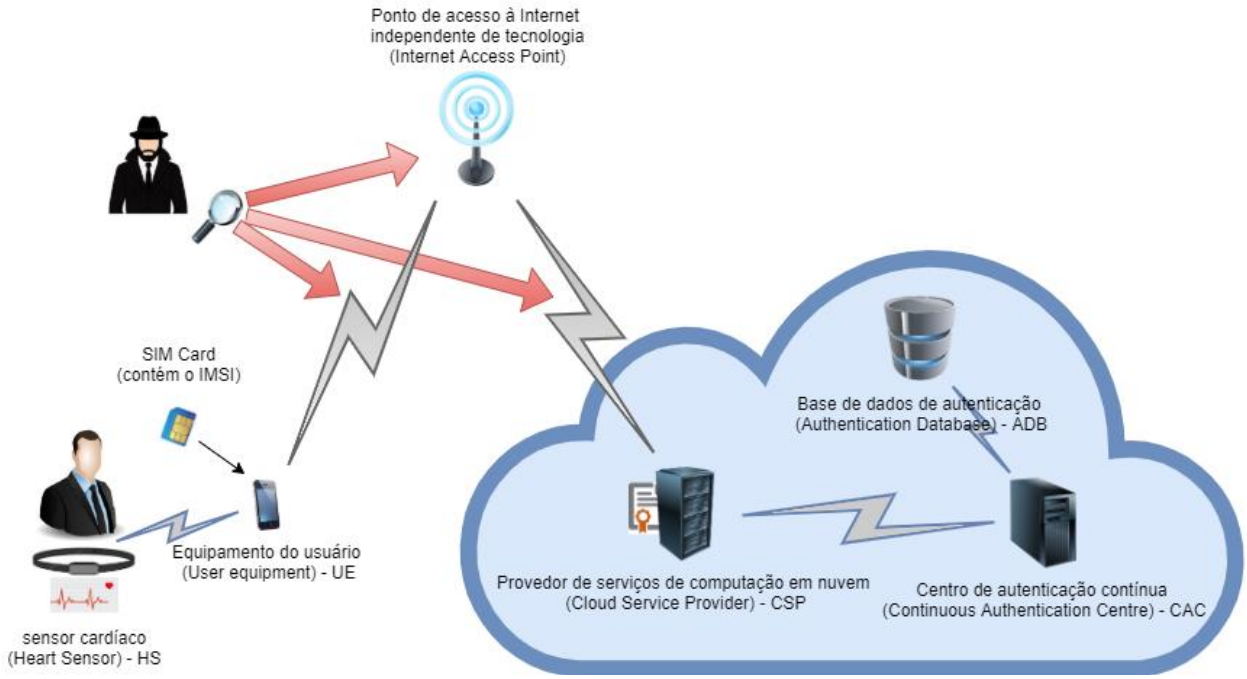

Figura 6.1 - Modelo de rede considerado

Considerando que o protocolo proposto visa a uma autenticação "fim a fim", o mesmo é concebido para funcionar na camada de aplicação. Dessa forma, os diversos aspectos envolvidos nas comunicações entre as várias entidades componentes não serão alvo de análise. É possível, por exemplo, que o HS se comunique com o UE por meio de uma conexão *Bluetooth*, o UE se comunique com o CSP por meio de uma conexão à Internet via WLAN ou redes móveis celulares (3G, 4G, 5G, etc.) e as entidades internas da nuvem se comuniquem utilizando as tecnologias disponíveis na própria nuvem (conexões *ethernet* utilizando cabos diversos, fibras ópticas ou até mesmo utilizando comunicações sem fio).

Para o modelo indicado, os canais de comunicações considerados inseguros e que, por conseguinte, estão sujeitos à intervenção de possíveis atacantes ou intrusos, são aqueles existentes entre o UE e o CSP, conforme fica indicado na figura anterior na qual um

atacante em potencial está monitorando os canais indicados. Os outros ambientes (região próxima ao usuário na qual estão presentes o próprio indivíduo, o HS e o UE com o SIM *Card*, e ambiente interno da nuvem formado pelo CSP, pelo CAC e pela ADB) são considerados livres de ameaças.

6.3- DEFINIÇÃO DO MODELO DE AUTENTICAÇÃO

O Algoritmo 6.1 fornece uma visão geral do funcionamento do protocolo de autenticação que será detalhado nos itens seguintes.

Algoritmo 6.1 - Funcionamento geral do protocolo de autenticação proposto

```
INÍCIO
    Selecionar a Fase
    Se Fase = Registro Então
            Usuário fornece Conj={Senha, IMSI e Dados brutos de ECG }
                                                        para o CSP
            CSP repassa Conj para o CAC
            CAC extrai as features de ECG e treina o classificador
            CAC armazena na ADB os dados necessários para a autenticação
                                    do usuário (IMSI, hash da senha e
                                    parâmetros do classificador treinado)
    Se Fase = Autenticação Então
            Se for Inicial Então
                    Usuário fornece Conj={Senha, IMSI e Dados brutos de
                                                    ECG } para o CSP
                    CSP repassa Conj para o CAC
                    CAC extrai as features de ECG e autentica o usuário
                            usando a senha, o IMSI e o classificador treinado
            Se for Contínua Então
                    Usuário fornece Dados brutos de ECG para o CSP
                    CSP repassa Dados brutos de ECG para o CAC
                    CAC extrai as features de ECG e autentica o usuário
                                            usando o classificador treinado
    Se Fase = Atualização Então
            Se for em ambiente controlado Então
                    Usuário fornece Conj'={nova Senha, IMSI e Dados brutos
                                                    de ECG} para o CSP
                    CSP repassa Conj' para o CAC
```

CAC extrai as features de ECG e treina o classificador
CAC atualiza, na ADB, os dados necessários para a autenticação do usuário (hash da nova senha e parâmetros do classificador treinado)
Se for durante uma sessão autenticada Então
Usuário fornece Conj''={nova Senha e Dados brutos de ECG} para o CSP
CSP repassa Conj'' para o CAC
CAC extrai as features de ECG e treina o classificador
CAC atualiza, na ADB, os dados necessários para a autenticação do usuário (hash da nova senha e parâmetros do classificador treinado)
Fim da Seleção
FIM

6.3.1- Registro

Nessa fase, que é executada antes do protocolo de autenticação começar a, de fato, autenticar, os dados pertinentes (ECG, IMSI, senha) do usuário são capturados e enviados para a nuvem a fim de que esses sejam armazenados na ADB para futuros processos de autenticação. Esta fase deve ser desencadeada em um ambiente físico controlado no qual o CSP esteja seguro em relação à identidade do usuário.

No protocolo proposto, para o registro e para as atualizações, é prevista a captura contínua de dados brutos de ECG até que a quantidade mínima necessária de ciclos reconhecidos para o treinamento do classificador seja obtida e a fase de registro seja concluída. No Capítulo 7-"Avaliação da proposta e resultados" conclui-se que essa quantidade mínima de ciclos deve ser igual 300.

O diagrama de sequência da Figura 6.2 representa a fase de registro na qual os seguintes passos são desencadeados:

- O UE recebe uma senha (*Pwd*) cadastrada diretamente pelo usuário, o IMSI obtido

a partir do SIM *Card* e dados brutos de ECG extraídos do usuário por meio de HS;

- Utilizando o certificado digital do CSP, o UE cria uma conexão SSL entre ele e o CSP por meio da qual são enviados o IMSI, o *H*(*Pwd*) (*hash* da senha criada pelo usuário) e os dados brutos de ECG;
- O CSP recebe os dados por meio da conexão SSL e os repassa, em claro, para o CAC;
- O CAC verifica a existência, na ADB, do IMSI que está sendo incluído;
- Caso não seja encontrado o IMSI, o CAC processa os dados brutos de ECG reconhecendo os seus ciclos, detectando os pontos fiduciais e extraindo as *features* de ECG a serem utilizadas no processo de autenticação (neste momento, caso a quantidade de ciclos reconhecidos seja inferior ao número de ciclos padronizado para o treinamento do classificador, a necessidade de mais dados brutos de ECG é sinalizada para as entidades envolvidas - CSP, UE e HS - e as mesmas providenciam a complementação);
- As *features* de ECG são, então, utilizadas para o treinamento do classificador que, depois de treinado, fornece seus parâmetros de treinamento para que sejam armazenados na ADB e usados, posteriormente, nos processos de autenticação;
- Estando tudo certo até esse momento, o CAC realiza o registro dos dados na ADB incluindo o IMSI, o *H(Pwd)* e os parâmetros do classificador.

É importante frisar que qualquer evento que fuja dos processos representados nos diagramas representativos das fases (a existência prévia de um IMSI que está sendo registrado ou a falha no processo de treinamento do classificador, por exemplo) interromperá automaticamente a fase que está sendo desencadeada gerando a comunicação adequada para as partes envolvidas e o retorno ao contexto anterior à execução dessa fase.

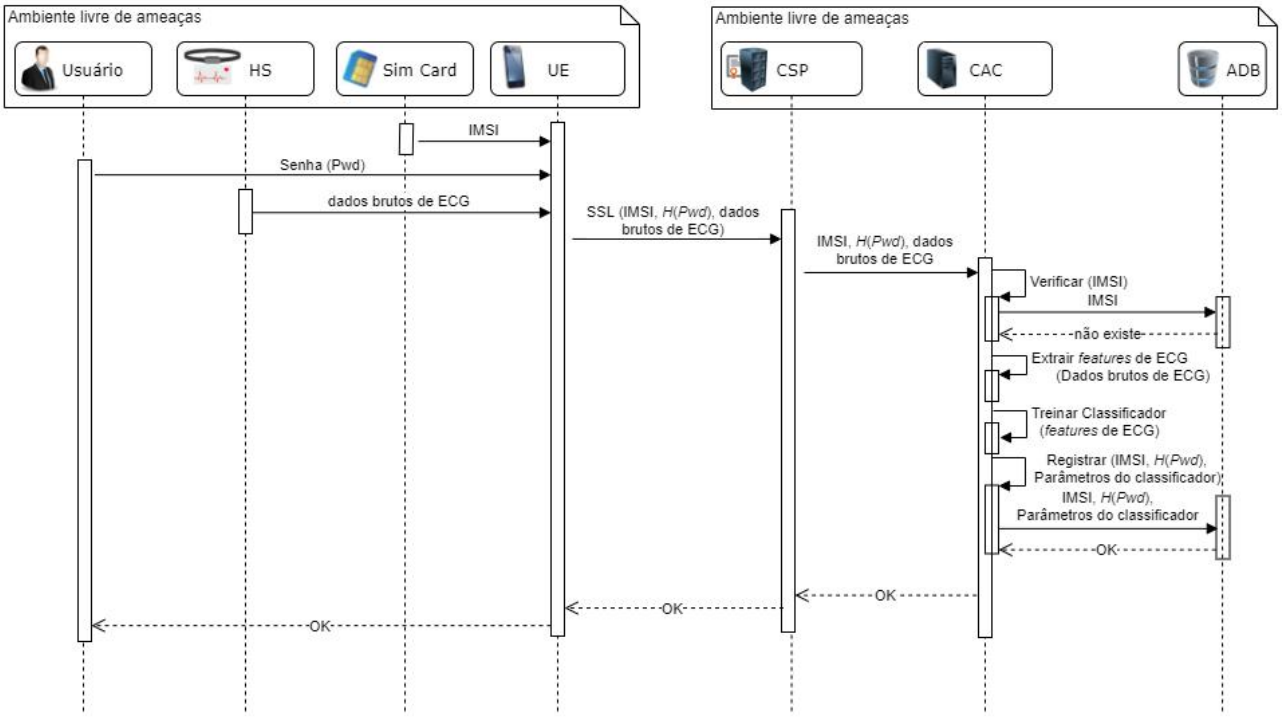

Figura 6.2 - Fase de registro

6.3.2- Autenticação

Fase que ocorre toda vez que um usuário deseja autentica-se perante o CSP a fim de solicitar algum serviço. Essa fase é divida em duas subfases:

- Autenticação inicial - na qual o usuário submete suas credenciais ao CSP a fim de ser autenticado no início de uma sessão de comunicação segura entre eles;
- Autenticação contínua - na qual o usuário, após a autenticação inicial e desejando permanecer autenticado perante o CSP, envia continuamente (de forma iterativa) seus dados brutos de ECG para o CAC.

Autenticação inicial

O diagrama de sequência da Figura 6.3 representa a subfase de autenticação inicial na qual os seguintes passos são desencadeados:

- O UE recebe a senha digitada diretamente pelo usuário, o IMSI obtido a partir do SIM *Card* e dados brutos de ECG extraídos do usuário por meio de sensores cardíacos;

- Utilizando o certificado digital do CSP, o UE cria uma conexão SSL entre ele e o CSP por meio da qual são enviados o IMSI, o *H(Pwd)* e os dados brutos de ECG;
- O CSP repassa para o CAC, em claro, os dados recebidos;
- O CAC verifica a existência, na ADB, do IMSI;
- Caso seja encontrado o IMSI, o CAC processa os dados brutos de ECG extraindo as *features* de ECG a serem utilizados no processo de autenticação;
- O CAC, então, autentica o usuário comparando o *H(Pwd)* recebido com *H(Pwd)* armazenado na ADB e classificando as *features* de ECG geradas utilizando o classificador treinado a partir dos parâmetros armazenados na base de autenticação;
- Caso a comparação de *hash* das senhas e a classificação indiquem que o usuário é quem diz ser, o CAC cria um identificador (*IdAut*) de autenticação da sessão de comunicação segura entre o CSP e o usuário;
- Após isso, o CAC passa o *IdAut* para o CSP que o repassa, via SSL, para o UE sinalizando que a autenticação inicial foi realizada com êxito e que o usuário já pode acessar os serviços a serem providos pelo CSP.

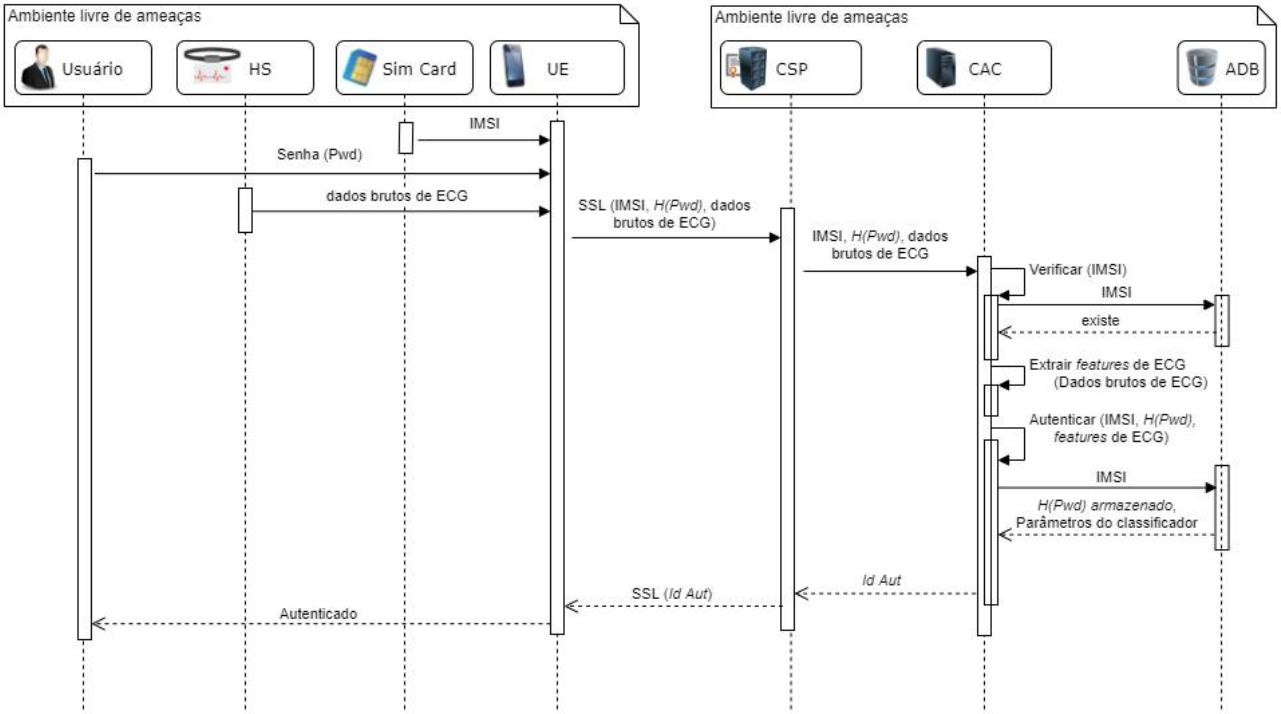

Figura 6.3 - Subfase de autenticação inicial

Cabe salientar que, tanto para a autenticação inicial quanto para a contínua, cada classificação realizada pelo algoritmo de ML sobre as *features* do ECG é realizada sobre dados equivalentes a um ciclo cardíaco verificado no eletrocardiograma. Dessa forma, os dados brutos de ECG extraídos do usuário devem ser em quantidade suficiente para que esse ciclo completo seja reconhecido.

Autenticação contínua

A subfase de autenticação contínua é executada, após a autenticação inicial, de forma iterativa, enquanto o usuário desejar permanecer autenticado perante o CSP. Nesse momento, já existe uma sessão de comunicação segura entre o CSP e o usuário identificada por *IdAut*.

O diagrama de sequência da Figura 6.4 representa uma iteração da subfase de autenticação contínua na qual os seguintes passos são dados:

- O UE recebe os dados brutos de ECG extraídos do usuário por meio de sensores cardíacos;
- Utilizando a conexão SSL já existente ente o UE e o CSP, são enviados o *IdAut* e os dados brutos de ECG;
- O CSP repassa para o CAC, em claro, os dados recebidos;
- O CAC processa os dados brutos de ECG extraindo as *features* de ECG a serem utilizadas no processo de autenticação;
- O CAC autentica o usuário por meio da classificação das *features* de ECG recém-obtidas com o classificador que já se encontra treinado;
- Tendo obtido êxito na autenticação, o CAC indica que a iteração de autenticação contínua foi realizada com êxito, devolve o *IdAut* para o CSP que o repassa ao usuário sinalizando que o mesmo pode permanecer acessando os serviços.

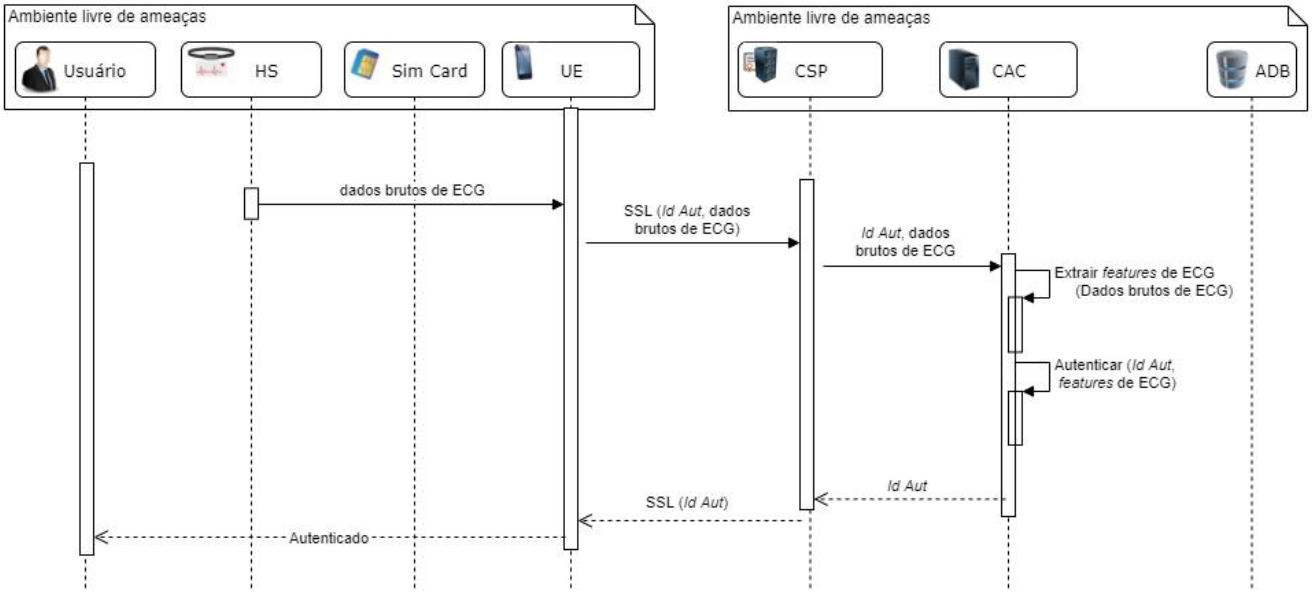

Figura 6.4 - iteração da subfase de autenticação contínua

É importante destacar que a autenticação contínua é realizada a cada ciclo cardíaco (cuja descrição é feita adiante, neste capítulo). Duas situações, nesse contexto, merecem atenção especial: o protocolo não reconhecer algum ciclo e não poder utilizá-lo para a autenticação, e o protocolo refutar a autenticação de algum ciclo. Considerando que o modelo proposto é multifatorial, ele possui um nível de segurança relevante independente da autenticação biométrica (há dois outros fatores). Assim, não é necessário que a sessão autenticada seja interrompida imediatamente após ocorrer uma das situações indicadas. Para esses casos, o protocolo considera as seguintes alternativas:

- Caso não seja reconhecido algum ciclo, novas tentativas são feitas sucessivamente após a falha e, permanecendo o problema por um intervalo de tempo superior a uma tolerância predefinida, a sessão é interrompida e seu reinicio é solicitado (o processo prossegue normalmente caso os ciclos voltem a ser reconhecidos antes do tempo da tolerância).
- Se algum ciclo for refutado na autenticação, de forma semelhante à situação anterior são feitas novas tentativas para tentar reverter o problema (se for revertido, o processo prossegue normalmente). Atingido um tempo máximo de tolerância, a sessão é interrompida, bloqueada e somente poderá ser desbloqueada após uma atualização em ambiente controlado (explicada a seguir).

6.3.3- Atualização

Fase que é executada quando alguma mudança precisa ser feita em um ou mais dados vinculados ao usuário que estão armazenados na ADB.

Esta fase pode ser motivada pelo esquecimento de uma senha, por uma mudança no padrão cardíaco (causada por uma cirurgia, por exemplo), pela necessidade (ou vontade) de alterar a senha ou treinar novamente o classificador. Cabe salientar que, como o IMSI é a base da identificação do usuário, esse dado não pode ser modificado na atualização. Para que o IMSI seja alterado, deverá ser desencadeada uma nova fase de registro do usuário.

A atualização pode ocorrer de duas formas:

- Atualização em ambiente controlado - ocorre quando o usuário não consegue mais se autenticar perante o CSP (esquecimento de senha, mudança no padrão cardíaco de autenticação, etc.) e deve ser desencadeada, semelhante à fase de registro, em um ambiente físico controlado no qual o CSP esteja seguro em relação à identidade do usuário;
- Atualização durante sessão autenticada - ocorre quando o usuário deseja alterar a sua senha e/ou treinar novamente o classificador. Essa atualização ocorre quando o usuário já se encontra autenticado e não necessita ser realizada em um ambiente físico controlado.

Atualização em ambiente controlado

O diagrama de sequência da Figura 6.5 representa a atualização em ambiente controlado na qual os seguintes passos são desencadeados:

- O UE recebe os dados a serem alterados (a senha nova - *Pwd* - e os dados brutos de ECG) e o IMSI;
- Utilizando o certificado digital do CSP, o UE cria uma conexão SSL entre ele e o CSP por meio da qual são enviados o IMSI, o *H(Pwd1)* e os dados brutos de ECG;
- O CSP repassa para o CAC, em claro, os dados recebidos;
- O CAC verifica a existência, na ADB, do IMSI que está sendo incluído;
- Caso seja encontrado o IMSI, o CAC processa os dados brutos de ECG extraindo as *features* de ECG a serem utilizadas no processo de autenticação (tal qual no

registro, mais dados devem ser providenciados caso não haja dados brutos de ECG suficientes para o treinamento do classificador);

- As *features* de ECG são, então, utilizadas para o treinamento do classificador que, depois de treinado, fornece seus parâmetros de treinamento para que sejam armazenados na ADB e usados, posteriormente, nos processos de autenticação;
- Estando tudo certo até esse momento, o CAC atualiza o registro dos dados na ADB modificando o *H(Pwd1)* e os parâmetros do classificador;
- Tendo obtido êxito na atualização, o CAC sinaliza para o CSP e para o usuário que a alteração foi realizada com êxito.

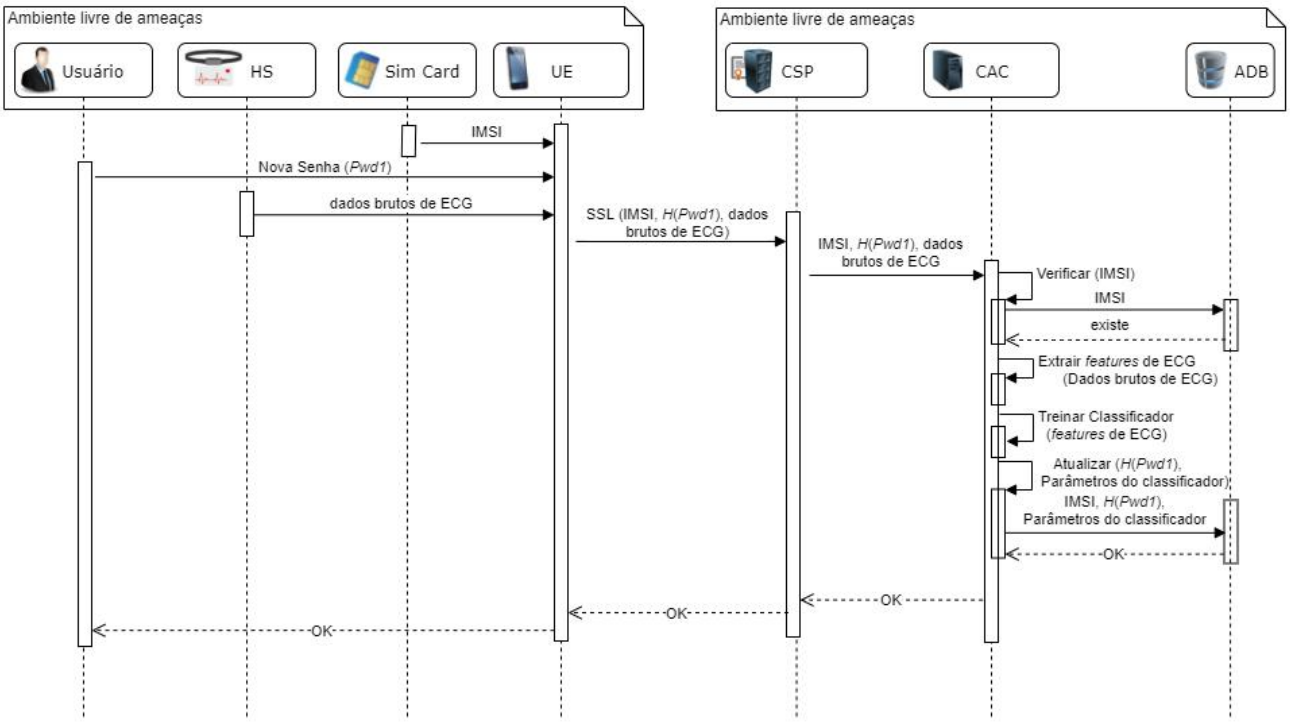

Figura 6.5 - Fase de atualização por meio de ambiente controlado

Atualização durante sessão autenticada

Como indicado anteriormente, essa atualização ocorre quando o usuário já se encontra autenticado perante o CSP. O diagrama de sequência da Figura 6.6 representa a atualização durante sessão autenticada na qual os seguintes passos são dados:

- O UE recebe os dados a serem alterados (a senha nova - *Pwd1* - e os dados brutos

de ECG);

- Utilizando a conexão SSL já existente ente o UE e o CSP, são enviados o *IdAut* o *H(Pwd1)* e os dados brutos de ECG;
- O CSP repassa para o CAC, em claro, os dados recebidos;
- O CAC processa os dados brutos de ECG extraindo as *features* de ECG a serem utilizadas no processo de autenticação (tal qual no registro e na atualização em ambiente controlado, mais dados devem ser providenciados caso não haja dados brutos de ECG suficientes para o treinamento do classificador);
- As *features* de ECG são, então, utilizadas para o treinamento do classificador que, depois de treinado, fornece seus parâmetros de treinamento para que sejam armazenados na ADB e usados, posteriormente, nos processos de autenticação;
- Estando tudo certo até esse momento, o CAC atualiza o registro dos dados na ADB modificando o *H(Pwd1)* e os parâmetros do classificador;
- Tendo obtido êxito na atualização, o CAC sinaliza para o CSP e para o usuário que a alteração foi realizada com êxito e prossegue com a sessão autenticada.

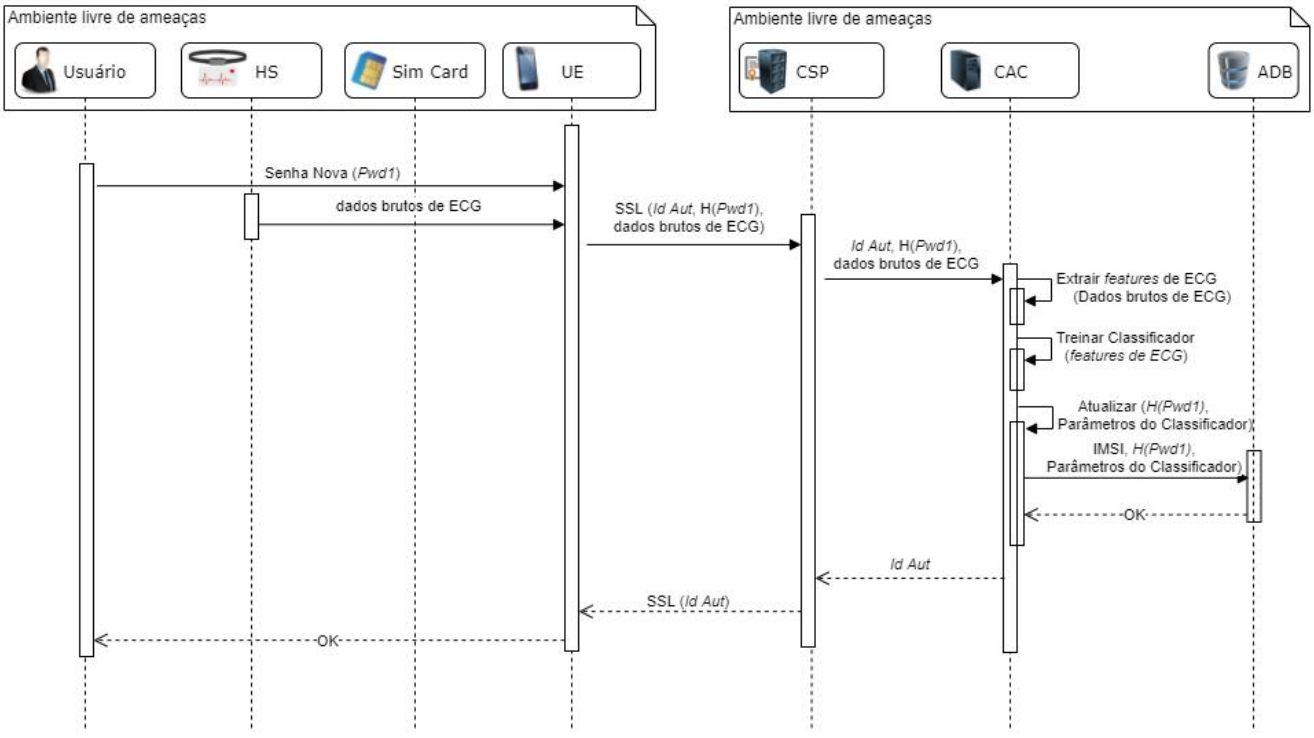

Figura 6.6 - Fase de atualização durante sessão autenticada

6.4- ASPECTOS ESPECÍFICOS DO PROCESSO BASEADO EM ELETROCARDIOGRAFIA

Antes de prosseguir com a descrição dos métodos que fazem parte do modelo de autenticação e que estão relacionados à eletrocardiografia, é importante salientar alguns aspectos:

- A presente proposta focaliza a autenticação de pessoas que não apresentam anomalias cardíacas e cujos ciclos cardíacos seguem um determinado padrão (indicado mais à frente). Essa premissa é utilizada principalmente para impor condições de detecção de pontos fiduciais utilizados na extração das *features*, como será visto nos próximos itens. Caso essa premissa não seja adotada, o algoritmo de detecção de pontos fiduciais utilizado deverá seguir outra lógica de funcionamento, algo que está fora do escopo deste trabalho.
- O ciclo cardíaco adotado nesta tese é uma curva que representa a parte de um ECG compreendida entre um ponto fiducial *LP* e o ponto fiducial *TP* subsequente, ou seja, é uma forma de onda que contem os pontos fiduciais *LP*, *P*, *Q*, *R*, *S*, *T* e *TP* que são a base para o processo de extração de *features* de ECG e que são representados na Figura 6.7. Esse ciclo deve seguir um padrão no qual, a partir de um ponto fiducial, o caminho para o ponto seguinte deve apresentar apenas um comportamento: ser ascendente ou descendente (podendo haver segmentos horizontais em parte do caminho).
- O **algoritmo de detecção de pontos fiduciais** que foi desenvolvido neste trabalho, apesar de estar baseado no padrão indicado, possui algumas faixas de tolerância que permitem o reconhecimento de um ciclo mesmo que ele apresente pequenas variações (configuráveis) no comportamento desejado.

Como pode ser percebido nas fases indicadas no item 6.3-"Definição do Modelo de Autenticação", os dados derivados de ECG constam de três métodos principais:

- Extrair *features* de ECG;
- Treinar classificador;
- Autenticar.

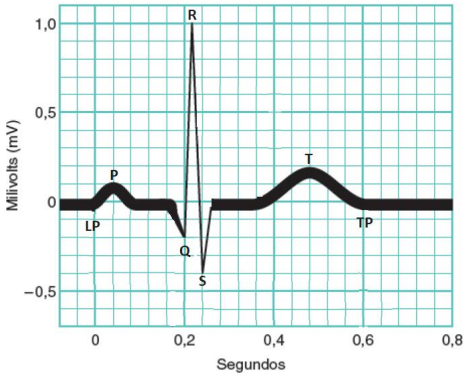

Figura 6.7 - Ciclo cardíaco típico
(adaptado de [TORTORA, 2014])

Todos esses métodos estão relacionados às *features* de ECG que são base para o emprego de ML no protocolo de autenticação proposto. O primeiro método é responsável por criar essas *features*, enquanto os outros dois as utilizam para promover a autenticação biométrica.

A seguir, cada um desses métodos será descrito focalizando o contexto da eletrocardiografia.

6.4.1- Método Extrair *features* de ECG

Nesse método é realizado um processamento sobre o sinal de ECG bruto (imagem fornecida pelo HS) a partir do qual são gerados dados para o treinamento e a classificação usando os algoritmos de ML. Esse processamento inclui **o algoritmo de detecção de pontos fiduciais** que foi desenvolvido para este propósito específico.

Cabe salientar que, no presente trabalho, todos os processamentos matemáticos foram realizados utilizando o *software* de processamento matemático ***MatLab*** em sua versão "R2016b".

A Figura 6.8 ilustra os passos desse método que são detalhados em seguida.

Filtragem

O vetor $E_0[t]$ representa os dados brutos de ECG e armazena valores de potenciais elétricos em mV (representam as amplitudes de um ECG), indexados temporalmente (a taxa de amostragem do ECG viabiliza uma relação entre um índice do vetor e um valor relativo de tempo). Esse vetor contém uma quantidade de ciclos determinada pela sua

duração e pela frequência cardíaca média do indivíduo. Como exemplo, caso $E_0[t]$ represente dois minutos de um ECG que apresenta frequência cardíaca média de 60 bpm, $E_0[t]$ conterá 120 ciclos.

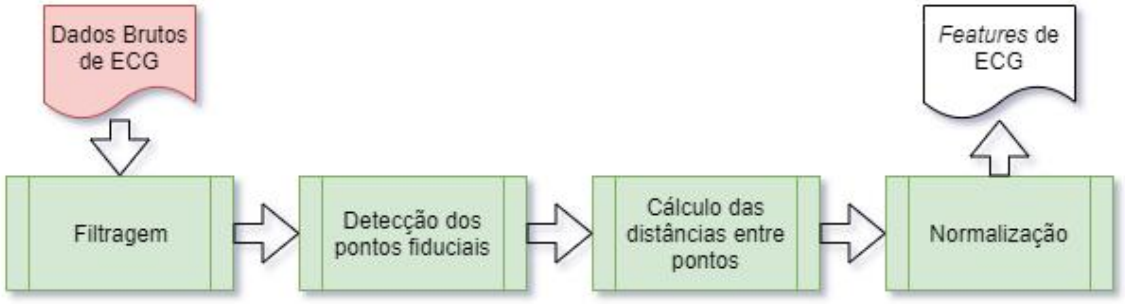

Figura 6.8 - Extração de *features* de ECG

Tendo, então, recebido um vetor $E_0[t]$, o sinal é processado para que possíveis ruídos que distorcem as posições exatas dos pontos fiduciais sejam atenuados. Para isso, é necessário utilizar-se um filtro passa-baixo (permite a passagem somente de baixas frequências). A opção deste trabalho foi pelo filtro *Butterworth* [MOREIRA, 2017] [RIVERA, 2018] que tem sido utilizado em alguns outros trabalhos relacionados à detecção de pontos fiduciais de ECG.

A aplicação do filtro dá-se por meio do uso das funções *butter* e *filter*, disponíveis no pacote de processamento de sinais do *software MatLab*, como pode ser visto no Trecho de código 6.1.

Trecho de código 6.1 - Utilização do filtro *Butterworth* no Matlab

```
data=E0; #vetor contendo dados brutos de ECG (base de dados ou captura com sensor)
[b,a]=butter (1, 60/500); #1ª ordem ; frequência de corte / taxa de amostragem ; 6/50 (valor da "sintonia")
E1 = filter(b,a,data);
```

O efeito da aplicação do filtro indicado pode ser percebido por meio da Figura 6.9, na qual se observa que os diversos picos e vales que representam o ruído do sinal são atenuados com a aplicação do filtro passa-baixo, o que permite uma detecção mais simples e precisa dos pontos fiduciais procurados.

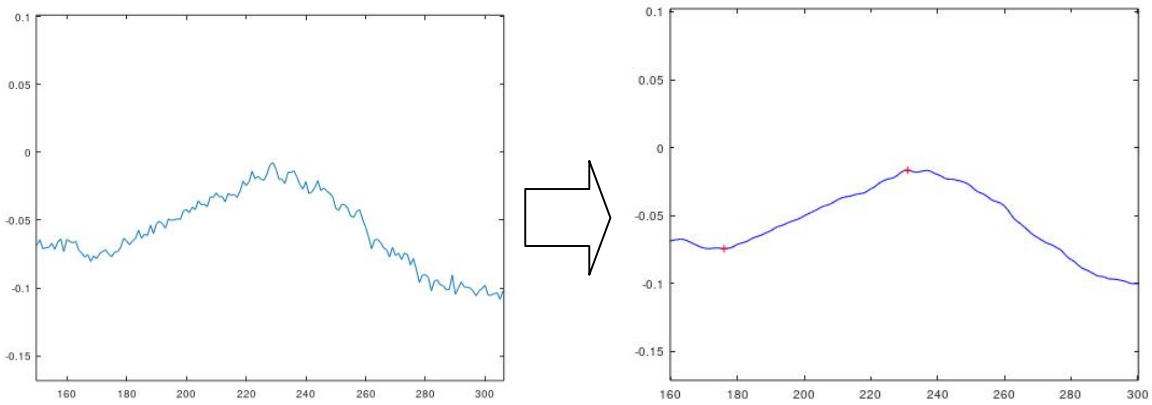

Figura 6.9 - Efeito da aplicação do filtro "Butterworth"

Após passar pelo processo de filtragem, o vetor $E_0[t]$ se transforma no vetor $E_1[t]$ que representa o sinal de ECG filtrado armazenando os seus valores de amplitudes.

Detecção dos pontos fiduciais

O algoritmo de detecção de pontos fiduciais implementado utiliza a função *findpeaks* do *MatLab* que indica uma série de picos e vales existentes no sinal analisado.

O Trecho de código 6.2 indica a forma de utilização da função sobre o sinal filtrado do ECG que se encontra armazenado no vetor $E_1[t]$.

Trecho de código 6.2 - Utilização da função *findpeaks*

```
[PKS IDX] = findpeaks(E1,"MinPeakDist",distH_min_picos);
# picos e vales ; distância horizontal = 0.02 s (valor obtido empiricamente)
```

A função é executada duas vezes, uma sobre o ECG normal (para a detecção dos picos) e outra sobre o ECG invertido (para a detecção dos vales). Em ambos os casos, é utilizado o valor de distância mínima entre picos de 0,02 s escolhido empiricamente a partir de experimentações.

Obtém-se, então, o vetor *IDX* que contém os valores dos índices de $E_1[t]$ que representam os instantes nos quais os picos e vales foram encontrados em $E_1[t]$. A Figura 6.10 indica, em azul, um ciclo filtrado, e em vermelho, os picos e vales encontrados.

Percebe-se que há determinados pontos fiduciais que são representados por mais de um pico/vale detectado (ponto *P* da figura indicada). Quando isso ocorre, o algoritmo de

detecção de pontos fiduciais assume o valor médio dos picos/vales que estão nas vizinhanças do ponto fiducial procurado. Caso o ponto esteja nas extremidades do ciclo analisado (pontos *LP* e *TP*), o pico/vale considerado será o mais interno (no caso de *LP*, o mais à direita, no caso de *TP*, o mais à esquerda).

Para finalizar a detecção dos pontos fiduciais, é realizada a escolha dos elementos adequados de *IDX* e são criados os vetores *LP[c], P[c], Q[c], R[c], S[c], T[c] e TP[c],* onde *LP*, *P*, ..., *TP* são vetores dos índices de $E_1[t]$ referentes a cada um dos pontos fiduciais detectados e *c* é o índice do ciclo cardíaco considerado.

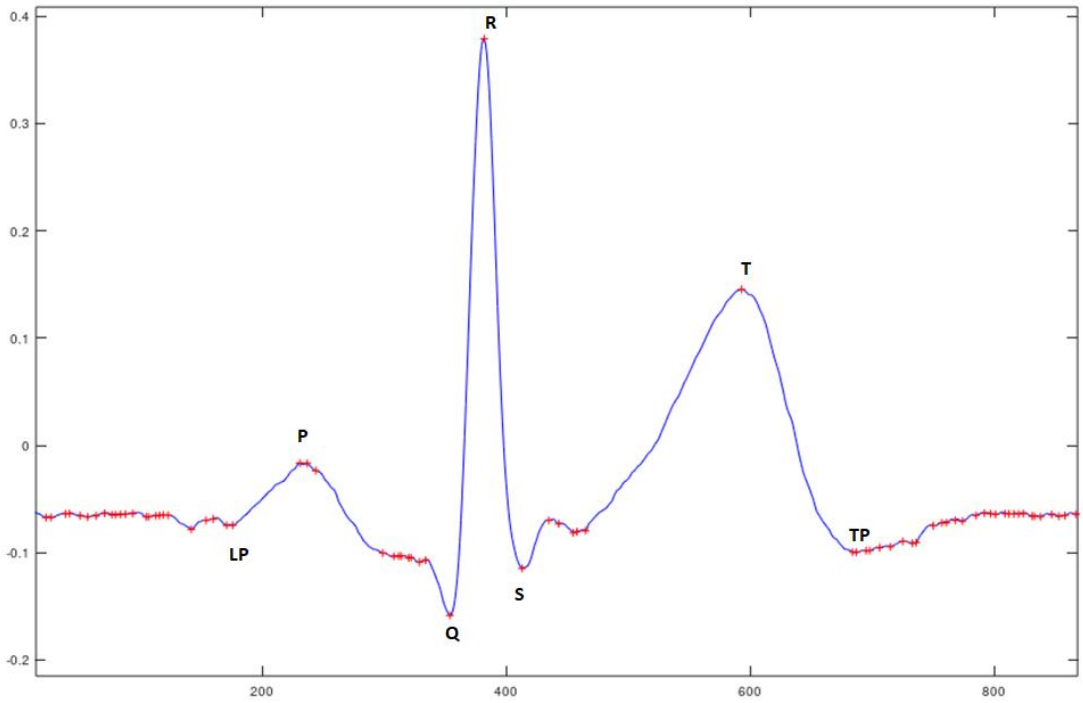

Figura 6.10 - Sinal filtrado de ECG com picos e vales detectados

Cálculo das distâncias entre pontos

Utilizando os vetores de índices *LP[c], P[c], Q[c], R[c], S[c], T[c] e TP[c],* e o vetor $E_1[t]$, que representa o sinal filtrado do ECG, é realizada a extração dos valores das distâncias representadas na Figura 6.11 por meio dos cálculos apresentados em seguida.

RLP(c) = R(c) – LP(c), onde *RLP(c)* representa a distância, em termos de índices, entre os pontos *R* e *LP* de *c;*

RP(c) = R(c) – P(c), análogo ao anterior;

RQ(c) = R(c) – Q(c), análogo ao anterior;

RS(c) = S(c) – R(c), análogo ao anterior;

RT(c) = T(c) – R(c), análogo ao anterior;

RTP(c) = TP(c) – R(c), análogo ao anterior;

RSA(c) = E_1(R(c)) - E_1(S(c)), onde *RSA(c)* representa a distância, em amplitude, entre os pontos *R* e *S* de *c;*

RQA (c) = E_1(R(c)) - E_1(Q(c)), análogo ao anterior.

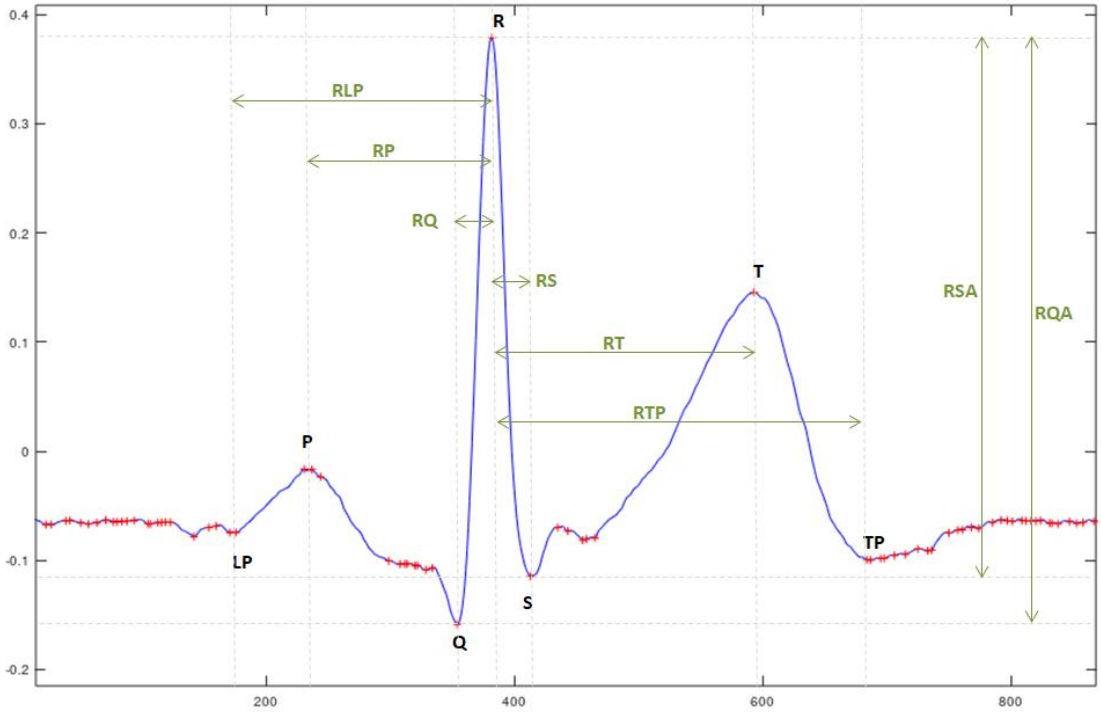

Figura 6.11 - Distâncias consideradas para o cálculo das *features* de ECG

Normalização

A Figura 6.12 representa uma amostra de 5 s tipicamente obtida a partir de um sinal de ECG que contém vários ciclos cardíacos que fornecerão, um a um, valores de pontos fiduciais a serem contabilizados para o cálculo final das *features* de ECG.

Caso os ciclos sejam separados e alinhados a partir de um mesmo ponto (geralmente do ponto *R*, que é o ponto fiducial mais destacado de um ECG normal), pode-se perceber que os comprimentos totais variam de ciclo para ciclo. Isso se deve às alterações da frequência cardíaca que, à medida que aumenta ou diminui, comprime ou expande os comprimentos dos ciclos, respectivamente.

Essas diferenças nos comprimento dos ciclos (e, por conseguinte, nos comprimentos dos segmentos internos aos ciclos) podem gerar distorções para os treinamentos dos classificadores, pois indicam medidas diferentes para a mesma característica do ciclo cardíaco.

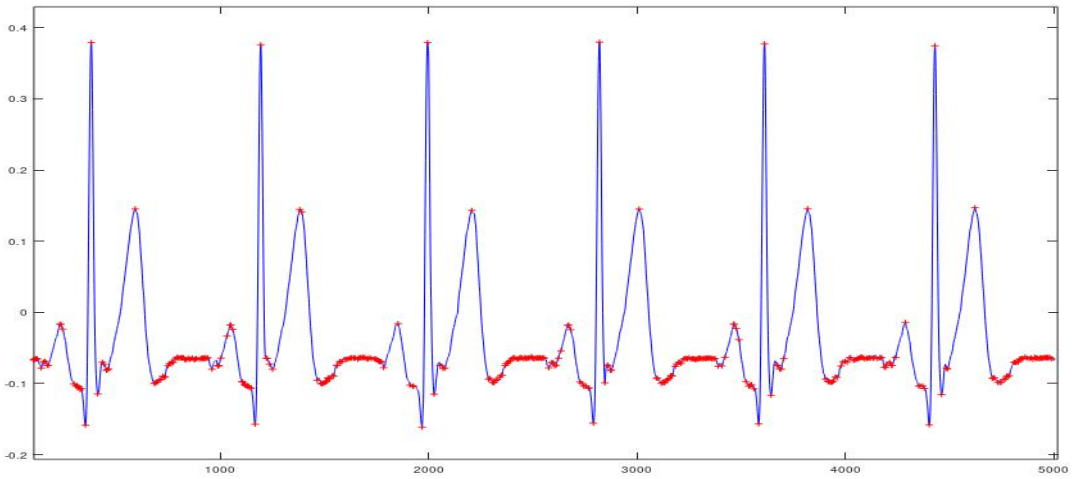

Figura 6.12 - Amostra de sinal de ECG de 5 s

A Figura 6.13 indica graficamente as diferenças entre os diversos comprimentos totais dos ciclos da amostra da Figura 6.12.

Para atenuar, então, os efeitos das variações da frequência cardíaca que geram diferentes comprimentos de ciclo, deve ser realizada uma normalização horizontal dos diversos ciclos analisados. Essa normalização é realizada por meio do cálculo da razão das distâncias encontradas anteriormente (em termos de índices) pelo comprimento total do ciclo (distância do ponto *LP* ao ponto *TP*, em termos de índices, de cada ciclo, ou simplesmente, distância *LP-TP*).

Outros trabalhos utilizam tipicamente as distâncias entre pontos *R* consecutivos (*R-R*) para a normalização, entretanto, como indicado no item 5.2.2-"Autenticação para equipamentos móveis baseada em ECG", a distância *R-R* não é muito adequada.

Nesta tese, a adoção da distância *LP-TP* para a normalização teve o principal objetivo de desprezar os efeitos da variação da frequência cardíaca que são percebidos na porção do ECG compreendida entre o ponto *TP* de um ciclo e o ponto *LP* do ciclo seguinte (ou seja, um trecho do ECG que está fora do ciclo cardíaco padronizado nesta tese). Apesar de não terem sido encontradas referências que corroborem o emprego dessa medida para esse fim, os resultados experimentais apresentados no próximo capítulo indicam que essa escolha é favorável.

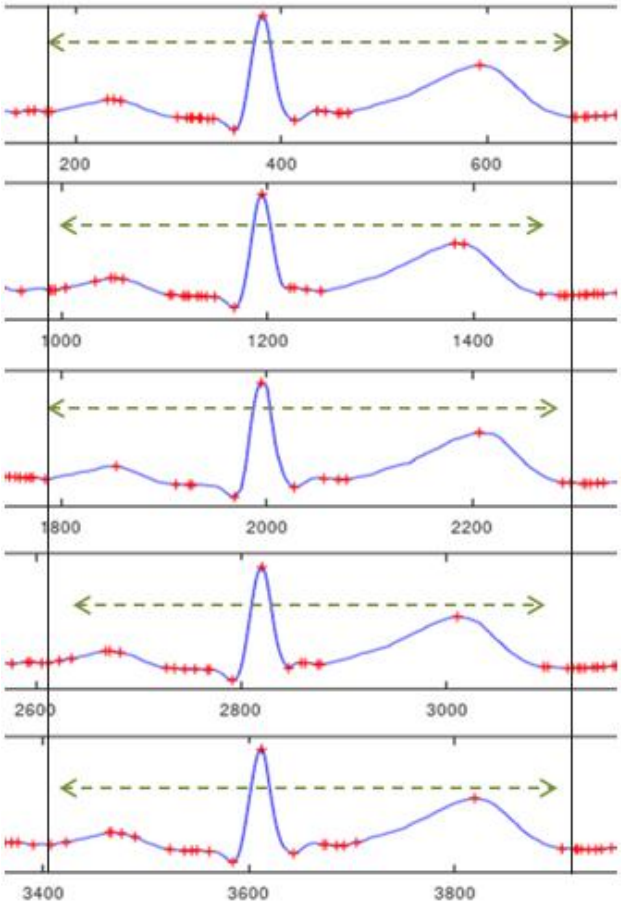

Figura 6.13. Diferenças dos comprimentos dos ciclos

Os cálculos da normalização finalizam, então, o processo de extração das *features* de ECG conforme as equações

$$f_1(c) = \frac{RLP(c)}{TP(c) - LP(c)},\ f_2(c) = \frac{RP(c)}{TP(c) - LP(c)},\ f_3(c) = \frac{RQ(c)}{TP(c) - LP(c)},$$

$$f_4(c) = \frac{RS(c)}{TP(c) - LP(c)},\ f_5(c) = \frac{RT(c)}{TP(c) - LP(c)},\ f_6(c) = \frac{RTP(c)}{TP(c) - LP(c)},$$

que representam as razões entre os segmentos e a maior distância do ciclo, como explicado anteriormente.

As duas outras *features* são dadas em mV (amplitude do ECG) e não passam pela normalização horizontal. As equações que as representam são

$$f_7(c) = RSA(c)\ e\ f_8(c) = RQA\ (c).$$

Os vetores $\boldsymbol{f_1[c], f_2[c], \ldots, f_8[c]}$ representam, finalmente, as *features* extraídas para todos os ciclos considerados.

Cabe salientar que, apesar dessas *features* serem derivadas das mesmas ideias básicas que as obtidas em [FALCONI, 2016], as suas formas de obtenção diferem nos contextos da detecção dos pontos fiduciais e da normalização.

6.4.2- Método Treinar classificador

No presente trabalho, são empregados os quatro classificadores indicados no capítulo 4. Para cada classificador é empregada uma forma específica de treinamento, como indicado a seguir.

Treinamento da SVM, do AdaBoost e do Robust Boost

Para os três primeiros classificadores (SVM, *AdaBoost* e *Robust Boost*), são utilizados algoritmos já existentes no *MatLab*.

O treinamento do classificador SVM consiste exclusivamente em utilizar a função *svmtrain* e passar, como exemplos para o treinamento, uma matriz de valores de *features*, como será visto detalhadamente no próximo capítulo.

Após o treinamento, a função apontada retorna uma estrutura de dados (*svmstruct*) própria do *MatLab* que contém todos os parâmetros desse classificador que devem ser armazenados na ADB.

Para o *AdaBoost* e o *Robust Boost*, é utilizada a função *fitcensemble* do *MatLab* passando, como parâmetro de escolha do tipo de classificador, os valores *AdaBoostM1* ou *RobustBoost*, e como exemplos para o treinamento, a mesma matriz de valores de *features* utilizada para a SVM (detalhada no capítulo 7).

Após o treinamento, de maneira análoga à SVM, a função retorna estruturas de dados que contêm todos os parâmetros dos classificadores que precisam ser guardados na ADB.

Treinamento do NNS

Para esse classificador, não são utilizadas funções já existentes no *MatLab*. O treinamento do classificador consiste exclusivamente em determinar os valores das médias das *features* por todos os ciclos considerados na análise, ou seja,

$$\mu_{f1} = \frac{1}{NC}\sum_{c=1}^{NC} f_1\ (c),\ \ \mu_{f2} = \frac{1}{NC}\sum_{c=1}^{NC} f_2\ (c),$$
$$\vdots$$
$$\mu_{f7} = \frac{1}{NC}\sum_{c=1}^{NC} f_7\ (c),\ \ \mu_{f8} = \frac{1}{NC}\sum_{c=1}^{NC} f_8\ (c),$$

onde μ_{fj} é a média da "j-ésima" *feature* considerando todos os ciclos do treinamento (*NC*).

Após os cálculos, os valores das médias são armazenados na ADB como parâmetros do classificador.

6.4.3- Método Autenticar

De forma semelhante ao método "Treinar", o método "Autenticar" também faz uso de funções do *MatLab*.

Autenticação com a SVM, o AdaBoost e o Robust Boost

A autenticação por meio desses classificadores baseia-se na utilização das funções *svmclassify* (para a SVM) e *predict* (para o *AdaBoost* e o *Robust Boost*). Em ambos os casos são passados, para as funções, os parâmetros dos classificadores obtidos da ADB e um vetor de *features* para a autenticação (lembrando que, neste momento, cada ciclo cardíaco representado por 8 *features* será autenticado isoladamente).

As funções indicadas, após a classificação, retornam um resultado positivo ou negativo para sinalizar a autenticação.

Autenticação com o NNS

Semelhante ao treinamento, para esse classificador não são utilizadas funções já existentes no *MatLab* para a classificação. Nesse caso, o processo é realizado por meio do Algoritmo 6.2.

Algoritmo 6.2 - Autenticação baseada em NNS

INÍCIO

Acertos = 0

Para cada feature fj baseada em tempo (ou seja, f1 a f6), fazer:

Se $(\mu_{fj} - Tolj) \leq fj \leq (\mu_{fj} + Tolj)$ *então Acertos = Acertos + 1*

Se (Acertos ≥ AcertosMin) E

$(\mu_{f7} - Tol7) \leq f7 \leq (\mu_{f7} + Tol7)$ *E* $(\mu_{f8} - Tol8) \leq f8 \leq (\mu_{f8} + Tol8)$

Então

AUTENTICADO!

Senão

NÃO AUTENTICADO!

FIM

Como o algoritmo indica, todas as *features* f_j baseadas em tempo (distâncias horizontais do ECG) são comparados às médias μ_{fj} armazenadas na ADB com base em tolerâncias específicas Tol_j que foram obtidas empiricamente. O total de êxitos nas comparações será registrado na variável *Acertos*. Para que o usuário seja considerado autenticado, a quantidade de acertos deve ser maior ou igual a um valor mínimo

AcertosMin e as duas *features* baseadas em amplitude (f_7 e f_8) devem estar dentro da faixa de valores adequada (considerando as médias armazenadas e as tolerâncias).

6.5- CONCLUSÃO

O presente capítulo tem o objetivo de apresentar a proposta central desta tese: um protocolo de autenticação multimodal, contínua, baseada em eletrocardiografia e em aprendizado de máquina, dentre outros aspectos. Esse protocolo tem sua aplicação focalizada em um ambiente de MCC cujos problemas de segurança foram explorados nos capítulos inicias deste trabalho.

São apresentadas algumas considerações iniciais seguidas dos modelos de rede e de autenticação. Este modelo, especificamente, tem as suas três fases descritas de forma bastante detalhada. Por fim, são tratados aspectos específicos do processo baseado em eletrocardiografia que permitem a compreensão, de forma precisa, dos processos de extração de *features* de ECG, de treinamento dos classificadores baseados em ML e de classificação com base nessas técnicas de aprendizado de máquina.

Como conclusão obtida a partir da análise dos diversos processos detalhados, pode-se afirmar que a proposta, ao abranger os vários aspectos necessários ao entendimento correto do protocolo, viabiliza a percepção de que ele apresenta várias características desejáveis para a resolução de um problema de autenticação baseada em biometria.

7- AVALIAÇÃO DA PROPOSTA E RESULTADOS

7.1- VALIDAÇÃO DO PROCESSO DE AUTENTICAÇÃO BASEADO EM ELETROCARDIOGRAFIA UTILIZANDO BASE DE ECG PÚBLICA

Com o propósito de validar o processo de autenticação baseado em eletrocardiografia, são realizadas algumas atividades que se encontram descritas nos itens a seguir.

Todos os cálculos e algoritmos executados no contexto deste capítulo foram realizados utilizando-se o *software MatLab* em sua versão R2016b. Além disso, o *hardware* empregado se resumiu a um computador pessoal com as seguintes configurações básicas: processador Intel i5, 16 GB de memória principal e 1 TB de capacidade de armazenamento em disco.

A Tabela 7.1 apresenta as variáveis e constantes utilizadas neste capítulo.

7.1.1- Bases de dados de Eletrocardiogramas

Foi escolhida, inicialmente, uma base de dados de ECG pública viabilizada pelo MIT, a *"MIT-BIH Arrhythmia Database"* [MIT-BIH, 2000a]. Considerando que o propósito principal do modelo proposto é autenticar usuários sem problemas cardíacos (cujos ECG não apresentam anomalias) por meio da detecção de padrões em ECG, essa escolha mostrou-se problemática pelo fato dessa base ser composta, em grande parte, por ECG representativos de pessoas com arritmia cardíaca, o que gera uma grande variabilidade dos tipos de forma de onda e que dificulta a análise dos ciclos cardíacos e o conseguinte treinamento inicial dos classificadores. Pelo fato da análise de ECG com anomalias não ser alvo deste trabalho, essa base foi descartada.

Partiu-se, então, para a busca de outras bases de dados de ECG e, após algumas pesquisas, optou-se pela utilização do repositório indicado em [SOUZA, 2019], que não focaliza usuários com anomalias cardíacas (apesar de não ter sido descartada, pelo autor, a participação desses usuários para a criação dessa base).

Tabela 7.1 - Variáveis e constantes utilizadas neste capítulo

Variável/ constante	**Descrição**	**Valor**
Nu_{Tot}	Total útil de usuários da base de dados de ECG, ou seja, número de usuários constantes da base que forneceram os ECG utilizados neste trabalho.	108
u	Índice de um usuário considerado da base	$1 \leq u \leq Nu_{Tot}$
$NC_{u\,Max}$	Número máximo de ciclos utilizados de cada usuário	300
NC_u	Número de ciclos cardíacos reconhecidos por usuário	$3 \leq NC_u \leq NC_{u\,Max}$
NC_{Tot}	Número total de ciclos reconhecidos considerando Nu_{Tot}	18.896
NC_{Max}	Número máximo de ciclos que podem ser adotados considerando Nu_{Tot} e $NC_{u\,Max}$	32.400
i	Índice de um ciclo cardíaco de um usuário	$1 \leq i \leq NC_{u\,Max}$
$c_{i,u}$	"i-ésimo" ciclo cardíaco de um usuário	$[f_{1,i,u}, f_{2,i,u}, \ldots f_{8,i,u}]$
Nf_{Tot}	Número total de *features*, por ciclo, utilizadas neste trabalho	8
j	Índice de uma *feature* do $c_{i,u}$	$1 \leq j \leq Nf_{Tot}$
$f_{j,i,u}$	"j-ésima" *feature* de um $c_{i,u}$	-
$\Delta t_{ECG\,Min}$	Duração mínima dos ECG considerados	300 s
M_u	Matriz de $f_{j,i,u}$ por usuário	-
M	Matriz de $f_{j,i,u}$ de todos os usuários	-
V	Matriz coluna de todos *u* relacionada à *M*	-
$Conj_{Class}$	Conjunto formado pelos 4 classificadores utilizados neste trabalho	[*SVM, AdaBoost (Ada), Robust Boost (Rob), NNS*]
Class	Classificador utilizado em uma análise específica	$Class \in Cj_{Class}$
$Conj_{Metr}$	Conjunto formado pelas 4 métricas de análise utilizadas neste trabalho	[*Acu*, *Pre*, *Sen*, *F1s*]
Metr	Métrica utilizada em uma análise específica	$Metr \in Cj_{Metr}$
Ns_{Tot}	Número total de sessões de validação da técnica de subamostragem aleatória, por usuário	100
s	Índice da sessão de validação da técnica de subamostragem aleatória	$1 \leq s \leq Ns_{Tot}$
$SAM_{u,s}$	Subamostra aleatória da matriz *M* referente ao usuário *u* e à sessão *s*	-
k	Índice (que também representa o tamanho) do grupo de sessões de validação da técnica de subamostragem aleatória	$10 \leq k \leq Ns_{Tot}$, com passos de 10
$Gs_{k,u}$	"k-ésimo" grupo de sessões de validação da técnica de subamostragem aleatória aplicada sobre o usuário	$Gs_{k,u} \in \begin{bmatrix} 10 \text{ sessões}, 20 \ \ldots \\ \ldots, 100 \text{ sessões} \end{bmatrix}$

Cabe lembrar, neste momento, que ciclos cardíacos (*c*), para o escopo deste trabalho, são curvas que representam partes de um ECG compreendidas entre um ponto fiducial *LP* e o ponto fiducial *TP* subsequente, ou seja, são formas de onda que contêm os pontos *LP, P, Q, R, S, T e TP*. Após reconhecer essas curvas, o processo de extração de *features* de ECG detecta os pontos fiduciais de cada ciclo e extrai, destes, as 8 *features* ($f1, f2, \ldots, f8$) utilizadas no processo de autenticação.

Da base escolhida, em um instante inicial, são aproveitados os dados de ECG de 108 usuários conforme o detalhamento da Tabela 7.2.

Como indicado, o total útil de usuários considerados é menor que o total bruto pelo fato de alguns processos de captura terem apresentado falhas na leitura dos sinais dos sensores de ECG. Essas falhas inviabilizam o aproveitamento dos sinais de alguns indivíduos, conforme explicado pelo autor do trabalho acadêmico do qual fez parte a criação da base considerada [SOUZA, 2019].

Tabela 7.2 - Características da base de ECG utilizada neste trabalho

Característica	**valor**
Total bruto de usuários	132
Nu_{Tot}	108
Quantidade de ECG por usuário	1
Total útil de homens	56
Total útil de mulheres	52
Faixa de idade 1 (18 a 35 anos)	98
Faixa de idade 2 (36 a 66 anos)	10
$\Delta t_{ECG\ Min}$	300s
Taxa de amostragem	500 Hz
Frequência cardíaca (menor que 60)	6
Frequência cardíaca (de 60 a 120)	102
Frequência cardíaca (acima de 120)	0

Durante as análises iniciais realizadas neste trabalho, verifica-se que a qualidade dos ECG da base escolhida é bastante variável (algo motivado, provavelmente, por problemas cardíacos dos participantes ou por problemas no processo de captura dos sinais elétricos que, apesar de não inviabilizarem a captura do sinal como um todo, geram ECG com várias distorções), o que prejudica a aplicação do algoritmo de detecção de pontos fiduciais do método de extração das *features* (detalhado no item 6.4.1-Método Extrair *features* de ECG") fazendo com que vários desses pontos não sejam detectados e

obrigando que os ciclos aos quais eles pertencem sejam considerados não reconhecidos e descartados da análise.

Neste trabalho, está definido que, para cada ECG existente na base de dados (1 ECG por usuário), são aproveitados, no máximo, 300 ciclos cardíacos. Essa quantidade foi escolhida pelo fato de $\Delta t_{ECG\ Min}$ ser igual a 300 segundos e da quase totalidade das FC serem iguais ou superiores a 60 bpm (ou seja, 1 ciclo por segundo). Esse limite superior de 300 ciclos está definido para que as análises contemplem, preferencialmente, NC_u de mesmo valor ($NC_u = NC_{u\ Max}$, para os usuários com $FC \geq 60\ bpm$) ou de valores muito próximos (os 6 usuários da base com $FC < 60\ bpm$ teriam NC_u um pouco menor que $NC_{u\ Max}$). Apesar dessa definição, NC_u varia bastante com valores inferiores a 300 ($3 \leq NC_u \leq 300$) por conta do problema de não detecção de pontos fiduciais citado anteriormente.

A Figura 7.1 e a Tabela 7.3 indicam os números de ciclos reconhecidos (NC_u) por usuário da base e mostram que muitos usuários tiveram menos de 300 ciclos reconhecidos.

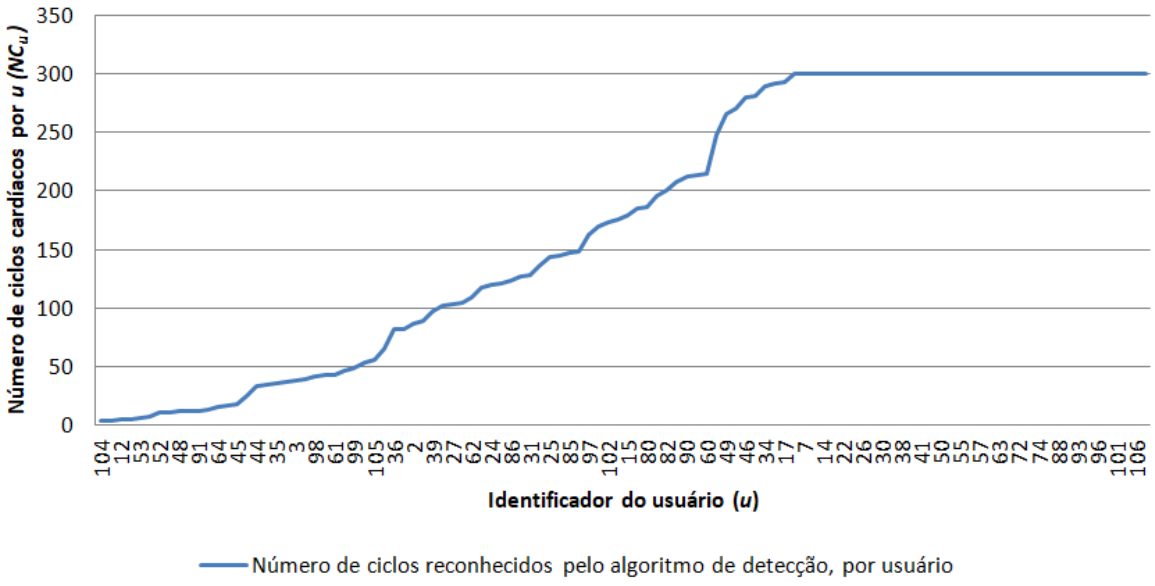

Figura 7.1 - Números de ciclos reconhecidos por usuário da base (ordenado por NC_u)

No presente trabalho, algumas tabelas seguem uma escala de cores baseada na ideia de mapa de calor. As cores dessa escala variam, em ordem crescente, do menor para o maior valor da coluna conforme a seguinte legenda:

Menores valores Maiores valores

Tabela 7.3 - Números de ciclos reconhecidos por usuário da base (ordenado por NC_u)

u	NC_u	u	NC_u	u	NC_u	u	NC_u	u	NC_u	u	NC_u	u	NC_u	u	NC_u
104	3	45	18	105	56	86	123	80	186	17	293	41	300	88	300
69	4	107	25	16	65	4	127	70	196	1	300	42	300	89	300
12	5	44	33	36	82	31	128	82	200	7	300	50	300	93	300
67	5	10	35	79	82	6	136	65	207	13	300	51	300	95	300
53	6	35	36	2	87	25	143	90	212	14	300	55	300	96	300
81	7	8	37	37	89	33	145	11	214	18	300	56	300	100	300
52	11	3	38	39	97	85	147	60	215	22	300	57	300	101	300
66	11	47	39	83	102	68	148	84	248	23	300	58	300	103	300
48	12	98	42	27	103	97	162	49	266	26	300	63	300	106	300
78	12	28	43	77	104	19	170	9	270	29	300	71	300	108	300
91	12	61	43	62	109	102	173	46	280	30	300	72	300		
76	13	94	46	87	117	21	175	20	281	32	300	73	300		
64	16	99	49	24	120	15	179	34	290	38	300	74	300		
43	17	59	53	54	121	5	185	92	292	40	300	75	300		

Como é possível perceber, a quantidade de ciclos reconhecidos por usuário (NC_u) varia bastante entre os usuários. Com o propósito de viabilizar uma percepção geral sobre os ciclos cardíacos utilizados, de fato, neste trabalho, a Tabela 7.4 indica alguns dados consolidados sobre essas quantidades de ciclos reconhecidos.

A Tabela 7.4 mostra, dentre outras informações, três faixas (ou conjuntos) de usuários separados tomando-se por base a quantidade de ciclos reconhecidos ($\boldsymbol{NC_u} = 300$, $25 < \boldsymbol{NC_u} < 300$, e $3 \leq \boldsymbol{NC_u} \leq 25$). Essa variação de quantidades de ciclos reconhecidos, por usuário, dificulta a avaliação dos resultados dos processos de autenticação, como é detalhado nos itens referentes à análise dos valores gerais das métricas. A faixa de usuários que apresenta mais problemas (dificuldade de encontrar padrões adequados, por exemplo, o que gera baixos valores das métricas de análise) referentes aos treinamentos e às classificações com os classificadores escolhidos é a dos usuários com menor quantidade de ciclos reconhecidos ($3 \leq \boldsymbol{NC_u} \leq 25$). A faixa seguinte

($25 < \boldsymbol{NC_u} < 300$), apesar de demonstrar melhores condições para a análise, também apresenta problemas. Já a melhor faixa ($\boldsymbol{NC_u} = 300$) apresenta poucos problemas.

Tabela 7.4 - Dados consolidados sobre as quantidades de ciclos reconhecidos

Característica	**valor**
NC_{Tot}	18.896
NC_{Max}	32.400
Média de ciclos reconhecidos por usuário $\left(\frac{NC_{Tot}}{Nu_{Tot}}\right)$	174,9
Usuários com $\boldsymbol{NC_u} = 300$	1 7 13 14 18 22 23 26 29 30 32 38 40 41 42 50 51 55 56 57 58 63 71 72 73 74 75 88 89 93 95 96 100 101 103 106 108 Total =37
Usuários com $25 < \boldsymbol{NC_u} < 300$	2 3 4 5 6 8 9 10 11 15 16 17 19 20 21 24 25 27 28 31 33 34 35 36 37 39 44 46 47 49 54 59 60 61 62 65 68 70 77 79 80 82 83 84 85 86 87 90 92 94 97 98 99 102 105 Total =36
Usuários com $3 \leq \boldsymbol{NC_u} \leq 25$	12 43 45 48 52 53 64 66 67 69 76 78 81 91 104 107 Total =16

7.1.2- Preparação das *features* utilizadas no modelo de autenticação

Uma vez concluída a detecção dos pontos fiduciais e o conseguinte reconhecimento dos ciclos cardíacos dos usuários, realiza-se a extração propriamente dita das *features* de todos os ciclos reconhecidos por usuário. Esses ciclos são representados por

$$c_{i,u} = \left[f_{1,i,u}, f_{2,i,u}, \ldots f_{8,i,u}\right] . \tag{7.1}$$

Da junção de todos os $c_{i,u}$ de um usuário obtêm-se as matrizes (M_u) de *features* de todos os ciclos reconhecidos por usuário. Essas matrizes possuem quantidades variáveis de linhas por 8 colunas (NC_u por Nf_{Tot}), e podem ser representadas por

$$M_u = \begin{bmatrix} f_{1,1,u} & f_{2,1,u} & \cdots & f_{8,1,u} \\ f_{1,2,u} & f_{2,2,u} & \cdots & f_{8,2,u} \\ & \ddots & & \\ f_{1,NC_u,u} & f_{2,NC_u,u} & \cdots & f_{8,NC_u,u} \end{bmatrix}_{NC_u \times Nf_{Tot}} .$$

(7.2)

Para a aplicação do processo de amostragem aleatória que será abordado à frente, as matrizes M_u são concatenadas criando a matriz *M* que possui 18.896 linhas por 8 colunas (NC_{Tot} por Nf_{Tot}), conforme representado em

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{Nu_{Tot}} \end{bmatrix}_{NC_{Tot} \times Nf_{Tot}} .$$

(7.3)

Além da matriz *M*, também é criada uma matriz coluna (*V*) que representa os números de identificação dos usuários que viabilizaram os ECG que originaram as *features* utilizadas. Essa matriz também possui 18.896 linhas que estão relacionadas diretamente às linhas de *M* de tal forma que as *features* de cada linha de *M* pertencem a um ciclo do ECG do usuário indicado na linha correspondente de *V*. Esta matriz coluna é representada por

$$V = \begin{bmatrix} 1 \\ \vdots \\ 108 \end{bmatrix}_{NC_{Tot} \times 1} .$$

(7.4)

Cabe salientar que, pelo fato de NC_u variar bastante (como dito anteriormente), a quantidade de linhas das matrizes *M* e *V* é inferior ao valor máximo possível. Caso todos os ciclos sejam reconhecidos, as matrizes indicadas terão, cada uma, 32.400 linhas ($\Delta t_{ECG\ Min} \times Nu_{Tot} = 300 \times 108 = 32.400$).

7.1.3- Processo de subamostragem aleatória para validação da autenticação

Após a criação das matrizes citadas anteriormente, adota-se uma técnica de subamostragem aleatória derivada da técnica de "espera" (do inglês, *holdout*) indicada em [KOHAVI, 1995] e composta por 100 sessões de validação por usuário. Essa técnica tem, como finalidade, validar o processo de autenticação baseada em eletrocardiografia utilizando as *features* obtidas durante o processo de extração. Um dos principais objetivos da técnica é verificar qual dos classificadores empregados (SVM, *AdaBoost*, *Robust Boost* e NNS) possui melhores condições para ser empregado no processo de autenticação.

A análise do processo de classificação considera quatro métricas definidas anteriormente neste trabalho: acurácia, precisão, sensibilidade e *F1-score*. Encontrando o melhor classificador com base nessas métricas, esse deve ser o escolhido para emprego no protocolo de autenticação.

As sessões de validação de cada usuário utilizam, cada, uma subamostra aleatória ($SAM_{u,s}$) das *features* da matriz *M* definida anteriormente. Cada $SAM_{u,s}$ é composta pelas seguintes matrizes:

- *M1_training* $_{u,s}$ - matriz obtida aleatoriamente de *M* contendo 70% dos ciclos de *u* (lembrando que os ciclos são formados pelas 8 *features* extraídas) e que serve para o treinamento dos classificadores;
- *M1_ validation* $_{u,s}$ - matriz obtida de *M* contendo os 30% restantes dos ciclos de *u* e que serve para a validação ou classificação feita pelos classificadores;
- *M0_training* $_{u,s}$ - matriz obtida aleatoriamente de *M* contendo 70% dos ciclos de usuários diferentes de *u* e que serve para o treinamento dos classificadores;
- *M0_ validation* $_{u,s}$ - matriz obtida aleatoriamente de *M*, tendo as mesmas dimensões de *M1_ validation* $_{u,s}$, sendo formada por ciclos de usuários diferentes de *u* e que serve para a validação ou classificação feita pelos classificadores.

As proporções "70% x 30%" aplicadas neste processo são arredondamentos das proporções "2/3 x 1/3" comumente encontradas na literatura relacionada a processos de treinamento e teste de classificadores (em [KOHAVI, 1995] e [REZGUI, 2016], por exemplo).

O Algoritmo 7.1 (representado por meio de pseudocódigo) ilustra os passos desencadeados durante o processo de subamostragem aleatória.

Algoritmo 7.1 - Passos da subamostragem aleatória

INÍCIO

Número total de sessões (Ns_{Tot}) = 100

Número total de usuários (Nu_{Tot}) = 108

Conjunto de classificadores ($Conj_{Class}$)= [SVM, Ada, Rob, NNS]

Conjunto de métricas ($Conj_{Metr}$)= [Acu, Pre, Sen, F1s]

Para cada usuário (u) de 1 até Nu_{Tot}, fazer:

Para cada sessão (s) de 1 até Ns_{Tot}, fazer:

Obter as matrizes M1_trainingu,s, M1_ validationu,s,M0_training u,s e M0_ validation u, s

Para cada classificador (Class) ∈ $Conj_{Class}$, fazer:

Treinar Class usando M1_trainingu,s e M0_training u,s

Para cada ciclo (c) ∈ [M1_ validationu,s ,M0_ validation u, s]

Classificar c como [verdadeiro_positivo (tp), falso_positivo (fp),verdadeiro_negativo (tn) ou falso_negativo (fn)]

Para cada métrica (Metr) ∈ $Conj_{Metr}$, fazer:

Calcular Metr usando as quantidades de tp, fp, tn e fn

FIM

De maneira resumida, o algoritmo acima realiza as seguintes tarefas

- Separação dos dados:
  - obtenção das matrizes *M1_training*$_{u,s}$, *M1_ validation*$_{u,s}$, *M0_training* $_{u,s}$ e *M0_ validation* $_{u,s}$;
- Treinamento dos classificadores:
  - para cada um dos 4 classificadores, é realizado o treinamento adequado com as matrizes obtidas anteriormente (*M1_training* e *M0_training*);
- Validação ou classificação:
  - de forma análoga ao treinamento, são realizadas as validações usando cada um dos classificadores já treinados e as *features* de *M1_validation* e *M0_validation*;

- Coleta dos resultados:
    - Para cada validação, é verificado se o processo "acerta" ou "erra", contabilizado-se o número total de acertos e erros para todas as amostras;
    - Os erros e acertos são definidos da seguinte maneira:
        - **verdadeiro positivo - TP** - classificador indica corretamente as *features* do padrão utilizado no seu treinamento;
        - **verdadeiro negativo - TN** - classificador indica corretamente as *features* que não são do padrão utilizado no seu treinamento;
        - **falso positivo - FP** - classificador indica erroneamente as *features* do padrão utilizado no seu treinamento;
        - **falso negativo - FN** - classificador indica erroneamente as *features* que não são do padrão utilizado no seu treinamento;
- Obtenção das métricas de análise:
    - Com base nos erros e acertos coletados, faz-se a contabilização das métricas de análise definidas anteriormente:
        - Acurácia;
        - Precisão;
        - Sensibilidade;
        - *F1-score*.

Verificação da "estabilidade" dos valores das métricas de análise

Após todas as sessões serem executadas por meio dos passos elencados anteriormente (100 sessões de subamostragens aleatórias por classificador por usuário), essas sessões são organizadas, para cada usuário, em 10 grupos ($Gs_{k,u}$) de sessões de validação da técnica de subamostragem aleatória. Esses $Gs_{k,u}$ têm quantidades que variam de 10 a 100 sessões, com passos de 10, ou seja,

$$Gs_{k,u} \in [10\ sessões, 20\ sessões, \dots, 100\ sessões] \quad . \tag{7.5}$$

O principal propósito dessa divisão em grupos de tamanhos variáveis é permitir a comparação dos valores médios obtidos de grupos cujos tamanhos aumentam gradativamente. Essa comparação viabiliza a verificação da "estabilidade" do valor médio das métricas (ou seja, a partir de um determinado tamanho de grupo, o valor médio não se altera de maneira relevante e pode ser adotado como valor médio da métrica no classificador considerado).

Feitos os agrupamentos indicados ($Gs_{k,u}$), são, então, calculados o valor médio μ e o desvio padrão σ de cada métrica de análise (*Acu*, *Pre*, *Sen* e *F1s*) de cada classificador (*SVM*, *Ada*, *Rob* e *NNS*) considerando cada $Gs_{k,u}$.

As equações (7.6) e (7.7) ilustram os cálculos genéricos de μ e σ:

$$\mu_{Class,Metr,k,u} = \frac{1}{k}\sum_{s=1}^{k} Metr_{Class,u,s} \tag{7.6}$$

e

$$\sigma\left(\mu_{Class,Metr,k,u}\right) = \sqrt{\frac{1}{k-1}\sum_{s=1}^{k}\left(Metr_{Class,u,s} - \mu_{Class,Metr,k,u}\right)^2}\,. \tag{7.7}$$

Como exemplo real de aplicação dessas equações genéricas, pode-se dizer que o valor médio da acurácia do classificador *SVM* aplicado sobre as *features* do usuário 1 em um grupo de 10 sessões de subamostragens aleatórias é calculado por

$$\mu_{SVM,Acu,10,1} = \frac{1}{10}\sum_{s=1}^{10} Acu_{SVM,1,s} \quad . \tag{7.8}$$

Da mesma forma, o desvio padrão relacionado ao valor médio calculado anteriormente pode ser dado por

$$\sigma\left(\mu_{SVM,Acu,10,1}\right) = \sqrt{\frac{1}{9}\sum_{s=1}^{10}\left(Acu_{SVM,1,s} - \mu_{SVM,Acu,10,1}\right)^2} \quad .$$

(7.9)

Calculados todos os valores médios e os desvios padrão das 17.280 combinações possíveis (4 classificadores, 4 métricas, 10 grupos de sessões e 108 usuários), faz-se, agora, o cálculo das médias desses μ e σ considerando todos os usuários, ou seja

$$\mu_{Class,Metr,k} = \frac{1}{Nu_{Tot}} \sum_{u=1}^{Nu_{Tot}} \mu_{Class,Metr,k,u} = \frac{1}{Nu_{Tot} \cdot k} \sum_{u=1}^{Nu_{Tot}} \sum_{s=1}^{k} Metr_{Class,u,s}$$

(7.10)

e

$$\mu(\sigma)_{Class,Metr,k} = \frac{1}{Nu_{Tot}} \sum_{u=1}^{Nu_{Tot}} \sigma\left(\mu_{Class,Metr,k,u}\right) =$$

$$= \frac{1}{Nu_{Tot} \cdot \sqrt{k-1}} \sum_{u=1}^{Nu_{Tot}} \sqrt{\sum_{s=1}^{k}\left(Metr_{Class,u,s} - \mu_{Class,Metr,k,u}\right)^2} \quad .$$

(7.11)

Exemplificando a aplicação real dessas equações genéricas, pode-se dizer que o valor médio (considerando todos os usuários) das médias da precisão do classificador *NNS* aplicado sobre as *features* de cada usuário em um grupo de 80 sessões de subamostragens aleatórias é calculado por

$$\mu_{NNS,Pre,80} = \frac{1}{Nu_{Tot} \cdot 80} \sum_{u=1}^{Nu_{Tot}} \sum_{s=1}^{80} Pre_{NNS,u,s} \quad .$$

(7.12)

Da mesma forma, o valor médio (considerando todos os usuários) dos desvios padrão relacionados às médias da sensibilidade do classificador *Ada* aplicado sobre as *features* de cada usuário em um grupo de 50 sessões de subamostragens aleatórias pode ser dado por

$$\mu(\sigma)_{Ada,Sen,50} = \frac{1}{Nu_{Tot} \cdot 7} \sum_{u=1}^{Nu_{Tot}} \sqrt{\sum_{s=1}^{50} \left(Sen_{Ada,u,s} - \mu_{Ada,Sen,50,u}\right)^2} \quad .$$

(7.13)

Concluídos todos os cálculos feitos com o propósito de verificar a estabilidade do valor médio das métricas, é possível indicar as Tabelas 7.5, 7.6, 7.7 e 7.8, além de traçar os gráficos das Figuras 7.2, 7.3, 7.4 e 7.5. Todos esses gráficos e tabelas indicam os resultados obtidos utilizando-se (7.10) e (7.11) e que representam os "limites de estabilidade" das métricas em cada classificador.

Tabela 7.5 - *SVM* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise

k	***Acu* ± *σ***	***Pre* ± *σ***	***Sen* ± *σ***	***F1s* ± *σ***
10	0,896 ± 0,041	0,888 ± 0,059	0,864 ± 0,062	0,868 ± 0,054
20	0,897 ± 0,040	0,888 ± 0,059	0,863 ± 0,064	0,868 ± 0,055
30	0,895 ± 0,040	0,886 ± 0,063	0,861 ± 0,064	0,866 ± 0,056
40	0,895 ± 0,040	0,887 ± 0,063	0,861 ± 0,064	0,866 ± 0,056
50	0,895 ± 0,039	0,887 ± 0,063	0,861 ± 0,064	0,866 ± 0,055
60	0,896 ± 0,039	0,887 ± 0,062	0,862 ± 0,063	0,866 ± 0,054
70	0,896 ± 0,039	0,886 ± 0,062	0,862 ± 0,064	0,866 ± 0,055
80	0,896 ± 0,039	0,887 ± 0,062	0,862 ± 0,064	0,867 ± 0,055
90	0,895 ± 0,039	0,886 ± 0,062	0,862 ± 0,064	0,866 ± 0,055
100	0,895 ± 0,039	0,886 ± 0,062	0,861 ± 0,063	0,866 ± 0,054

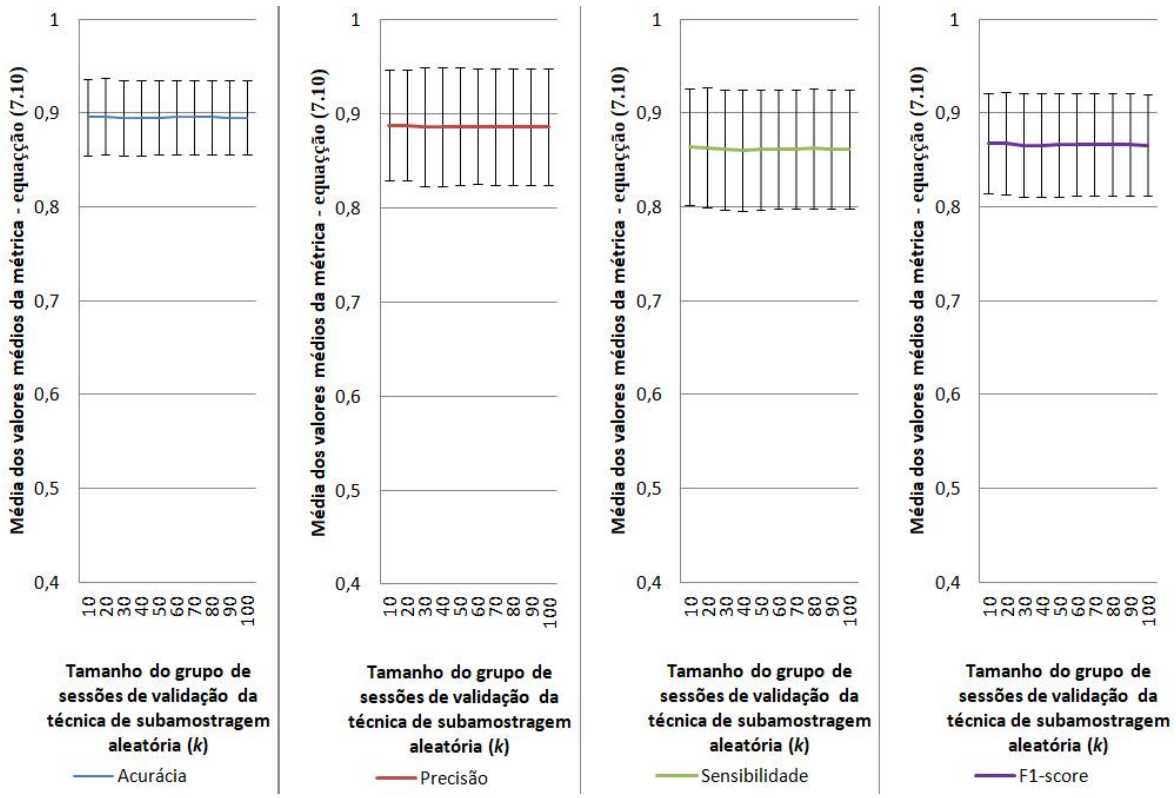

Figura 7.2 - *SVM* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise

Tabela 7.6 - *AdaBoost* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise

k	***Acu* ± *σ***			***Pre* ± *σ***			***Sen* ± *σ***			***F1s* ± *σ***		
10	0,808	±	0,039	0,845	±	0,061	0,617	±	0,078	0,692	±	0,073
20	0,808	±	0,040	0,840	±	0,055	0,617	±	0,079	0,691	±	0,073
30	0,807	±	0,039	0,840	±	0,055	0,615	±	0,079	0,689	±	0,073
40	0,807	±	0,039	0,841	±	0,055	0,615	±	0,079	0,689	±	0,073
50	0,807	±	0,040	0,841	±	0,056	0,614	±	0,080	0,689	±	0,074
60	0,807	±	0,040	0,841	±	0,056	0,615	±	0,079	0,690	±	0,074
70	0,807	±	0,040	0,842	±	0,055	0,615	±	0,079	0,690	±	0,073
80	0,807	±	0,040	0,842	±	0,056	0,615	±	0,079	0,690	±	0,074
90	0,807	±	0,040	0,842	±	0,056	0,615	±	0,079	0,690	±	0,074
100	0,807	±	0,040	0,842	±	0,057	0,615	±	0,079	0,690	±	0,074

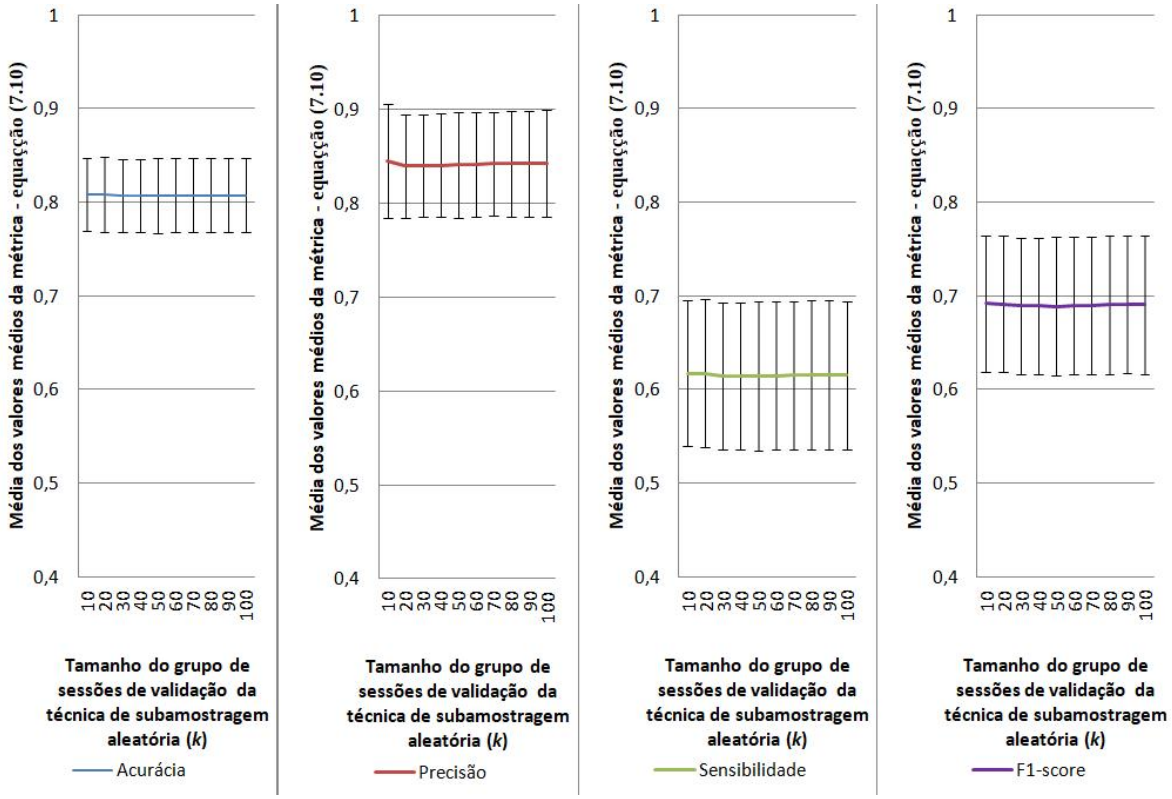

Figura 7.3 - *AdaBoost* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise

Tabela 7.7 - *Robust Boost* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise

k	***Acu* ± *σ***			***Pre* ± *σ***			***Sen* ± *σ***			***F1s* ± *σ***		
10	0,815	±	0,045	0,894	±	0,062	0,631	±	0,090	0,713	±	0,086
20	0,818	±	0,047	0,898	±	0,061	0,637	±	0,094	0,719	±	0,089
30	0,816	±	0,047	0,896	±	0,065	0,634	±	0,094	0,716	±	0,089
40	0,817	±	0,047	0,897	±	0,067	0,634	±	0,094	0,717	±	0,089
50	0,817	±	0,047	0,897	±	0,068	0,634	±	0,094	0,717	±	0,089
60	0,817	±	0,046	0,897	±	0,068	0,634	±	0,092	0,717	±	0,087
70	0,816	±	0,046	0,897	±	0,069	0,633	±	0,092	0,716	±	0,088
80	0,816	±	0,046	0,897	±	0,070	0,633	±	0,092	0,716	±	0,088
90	0,816	±	0,046	0,897	±	0,069	0,633	±	0,092	0,716	±	0,088
100	0,816	±	0,046	0,897	±	0,069	0,634	±	0,092	0,717	±	0,087

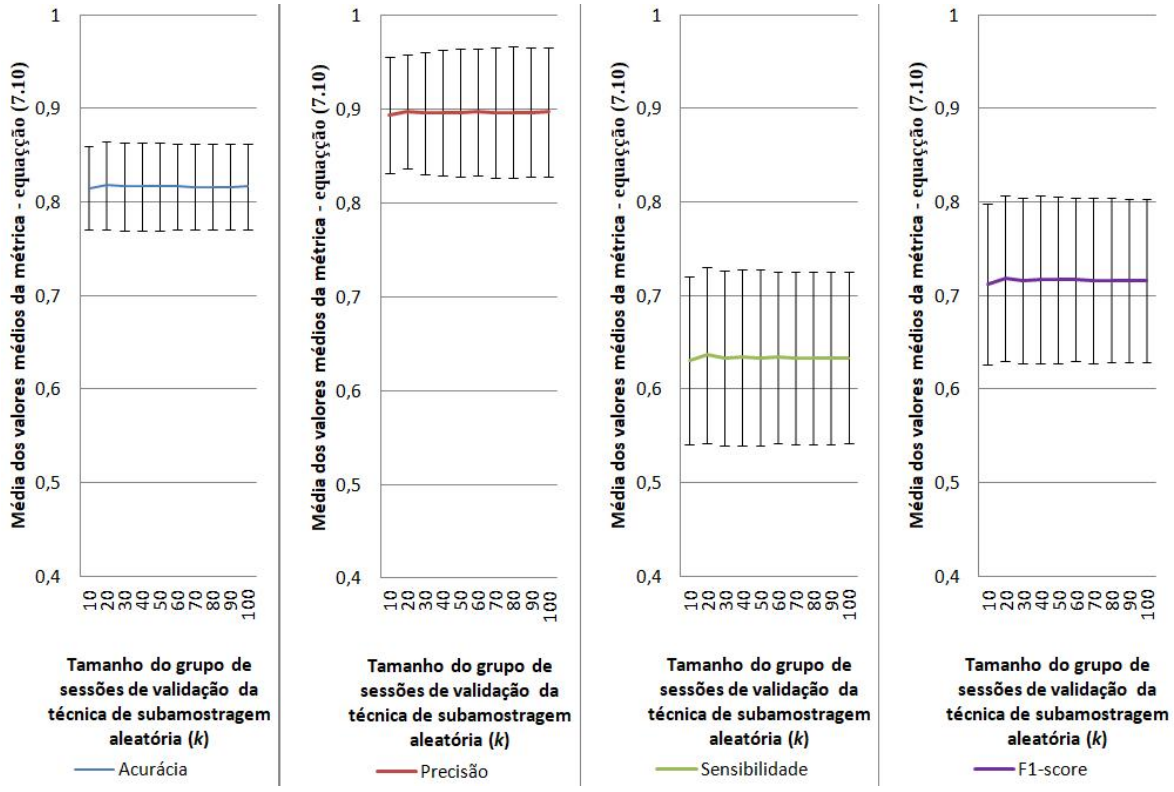

Figura 7.4 - *Robust Boost* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise

Tabela 7.8 - *NNS* - tabela que indica a estabilidade do comportamento dos valores médios das métricas de análise

k	***Acu* ± *σ***			***Pre* ± *σ***			***Sen* ± *σ***			***F1s* ± *σ***		
10	0,752	±	0,051	0,841	±	0,084	0,507	±	0,101	0,596	±	0,099
20	0,753	±	0,052	0,844	±	0,084	0,509	±	0,103	0,598	±	0,100
30	0,754	±	0,051	0,844	±	0,085	0,510	±	0,101	0,600	±	0,098
40	0,754	±	0,051	0,845	±	0,084	0,512	±	0,102	0,601	±	0,098
50	0,754	±	0,051	0,846	±	0,083	0,511	±	0,102	0,601	±	0,098
60	0,754	±	0,052	0,846	±	0,084	0,512	±	0,103	0,602	±	0,099
70	0,754	±	0,052	0,846	±	0,085	0,511	±	0,103	0,601	±	0,099
80	0,754	±	0,052	0,846	±	0,085	0,511	±	0,103	0,600	±	0,099
90	0,754	±	0,052	0,846	±	0,084	0,510	±	0,103	0,600	±	0,099
100	0,753	±	0,052	0,847	±	0,085	0,510	±	0,104	0,600	±	0,100

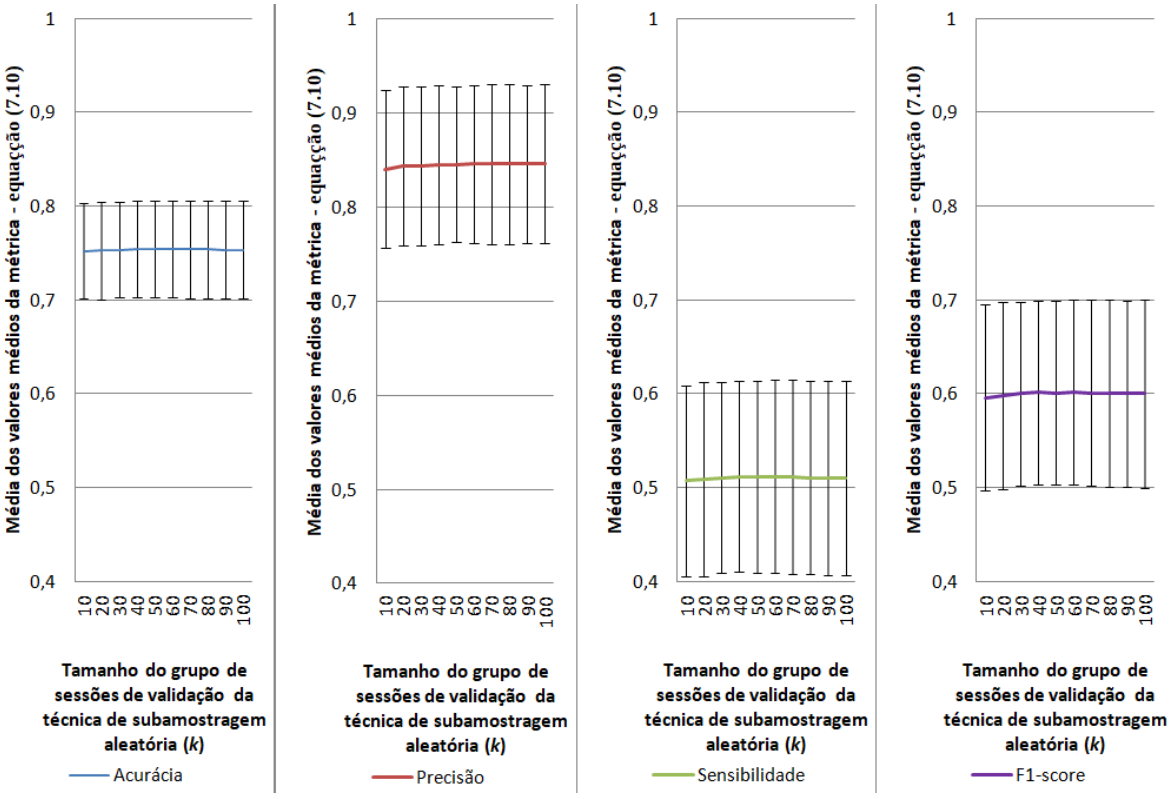

Figura 7.5 - *NNS* - gráficos que indicam a estabilidade do comportamento dos valores médios das métricas de análise

Para cada métrica, em cada classificador, é possível perceber, por meio dos gráficos apresentados, a estabilidade dos valores de suas médias observando o paralelismo de suas curvas (quase retas) em relação ao eixo das abscissas. Além disso, observando as tabelas, percebe-se que, em alguns casos, considerando a comparação entre médias relacionadas aos valores iniciais e finais de *k*, essa variação chega a 0,006, mas utilizando valores de *k* entre 30 e 100, essa variação cai para 0,001, o que sinaliza que esses valores de métricas não tendem a mudar muito com o aumento de *k*. Idealmente, quanto maior o valor dessa variável, mais representativos serão os valores das médias das métricas analisadas, entretanto avalia-se que a variação de 1 milésimo indicada anteriormente justifique a interrupção da evolução de *k* em 100.

Tendo em vista a estabilidade percebida anteriormente e o fato de que todas as métricas de todos os classificadores são calculadas em até 100 sessões (Ns_{Tot}), é válido supor que as médias das métricas calculadas com esse número máximo de sessões são, de fato, representações estabilizadas e podem ser adotadas como valores médios padrão para as métricas nos diversos classificadores.

Análise dos valores gerais das métricas

A partir da conclusão anterior referente à estabilidade, são adotadas, para as demais análises deste trabalho, somente as médias das métricas calculadas a partir do total de sessões realizadas ($k = Ns_{Tot}$). Por simplicidade, essas médias das métricas calculadas utilizando-se (7.10) com $k = Ns_{Tot}$ são chamadas, daqui em diante, somente de métricas (*Acu*, *Pre*, *Sen*, *F1s*) e, da mesma forma, as médias dos desvios padrão calculados por meio de (7.11) são chamadas somente de desvios padrão (σ), ou seja,

$$Metr_{Class} = \frac{1}{Nu_{Tot} \cdot Ns_{Tot}} \sum_{u=1}^{Nu_{Tot}} \sum_{s=1}^{Ns_{Tot}} Metr_{Class,u,s} \tag{7.14}$$

e

$$\sigma_{Class,Metr} = \frac{1}{Nu_{Tot} \cdot \sqrt{Ns_{Tot} - 1}} \sum_{u=1}^{Nu_{Tot}} \sqrt{\sum_{s=1}^{Ns_{Tot}} \left(Metr_{Class,u,s} - \mu_{Class,Metr,Ns_{Tot},u}\right)^2} \; .$$

(7.15)

Como resultados gerais, a Tabela 7.9, criada a partir de (7.14) e (7.15), indica, no contexto do presente trabalho, uma comparação geral das métricas e respectivos desvios padrão.

Tabela 7.9 - Mapa de calor representando os valores das métricas de análise e os desvios padrão em todos os classificadores utilizados .

Class	***Acu* ± *σ***			***Pre* ± *σ***			***Sen* ± *σ***			***F1s* ± *σ***		
SVM	0,895	±	0,039	0,886	±	0,062	0,861	±	0,063	0,866	±	0,054
Ada	0,807	±	0,04	0,842	±	0,057	0,615	±	0,079	0,69	±	0,074
Rob	0,816	±	0,046	0,897	±	0,069	0,634	±	0,092	0,717	±	0,087
NNS	0,753	±	0,052	0,847	±	0,085	0,51	±	0,104	0,6	±	0,1

Percebe-se que o classificador SVM apresenta os melhores valores de *Acu*, *Sen* e *F1s*, bem como dos seus respectivos desvios padrão. Além disso, em termos de *Pre* e seu respectivo *σ*, esse classificador posiciona-se em segundo lugar.

No outro extremo do desempenho, o classificador NNS apresenta os piores resultados de em 7 dos 8 valores observados. Apenas a *Pre* não segue essa regra, ficando em penúltimo lugar dentre as 4 opções.

Considerando os outros dois classificadores, merece destaque o *Rob*, que apresentou o maior valor de *Pre* e o segundo maior valor nas métricas *Acu*, *Sen* e *F1s*.

Buscando isolar os problemas que geraram reconhecimentos de ciclos inferiores ao máximo possível ($NC_u < NC_{Max}$), a Tabela 7.10 indica valores análogos aos da Tabela 7.9 usando somente os casos em que $NC_u = NC_{Max}$.

Tabela 7.10 - Mapa de calor análogo ao da Tabela 7.9 considerando apenas os casos em que $NC_u = NC_{Max}$.

Class	***Acu* ± σ**	***Pre* ± σ**	***Sen* ± σ**	***F1s* ± σ**
SVM	0,947 ± 0,015	0,928 ± 0,022	0,971 ± 0,016	0,949 ± 0,014
Ada	0,920 ± 0,032	0,962 ± 0,038	0,840 ± 0,064	0,893 ± 0,052
Rob	0,917 ± 0,030	0,983 ± 0,044	0,836 ± 0,060	0,895 ± 0,053
NNS	0,843 ± 0,033	0,995 ± 0,008	0,689 ± 0,065	0,791 ± 0,051

Apesar dos valores da Tabela 7.9 serem menores que os da Tabela 7.10, o principal resultado da comparação entre os classificadores, usando esta tabela, é muito próximo ao resultado obtido na comparação anterior, ou seja, o classificador SVM apresenta os melhores valores na grande maioria das métricas e dos desvios padrão analisados.

Apesar das conclusões anteriores representarem alguns dos principais resultados buscados por meio das análises deste capítulo, faz-se necessário o relato de algumas considerações referentes aos diversos classificadores aplicados sobre os vários usuários da base de dados escolhida. As análises que fundamentam essas considerações foram feitas com base nos dados das Tabelas 7.11, 7.12, 7.13 e 7.14, que contêm dados de alguns usuários, e do apêndice A, que contém figuras e tabelas com todos os usuários.

Tabela 7.11 - SVM - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,820 ± 0,241	0,640 ± 0,482	0,640 ± 0,482	0,640 ± 0,482
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,632 ± 0,137	0,603 ± 0,469	0,303 ± 0,264	0,390 ± 0,317
12	78	0,809 ± 0,109	0,990 ± 0,100	0,618 ± 0,217	0,739 ± 0,188
13	76	0,499 ± 0,013	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
16	64	0,922 ± 0,101	0,971 ± 0,075	0,874 ± 0,183	0,908 ± 0,131
25	107	0,841 ± 0,087	0,917 ± 0,106	0,766 ± 0,164	0,820 ± 0,109
36	35	0,807 ± 0,064	0,921 ± 0,085	0,680 ± 0,114	0,774 ± 0,085
39	47	0,928 ± 0,053	0,962 ± 0,056	0,896 ± 0,094	0,924 ± 0,059
42	98	0,723 ± 0,076	0,863 ± 0,118	0,534 ± 0,130	0,651 ± 0,116
43	61	0,910 ± 0,060	0,910 ± 0,079	0,919 ± 0,084	0,911 ± 0,061
53	59	0,903 ± 0,056	0,984 ± 0,033	0,821 ± 0,107	0,891 ± 0,069
82	36	0,925 ± 0,036	0,944 ± 0,047	0,907 ± 0,051	0,924 ± 0,037
89	37	0,835 ± 0,048	0,937 ± 0,051	0,720 ± 0,090	0,811 ± 0,063
103	27	0,952 ± 0,022	0,964 ± 0,033	0,941 ± 0,035	0,951 ± 0,022
117	87	0,820 ± 0,042	0,845 ± 0,047	0,787 ± 0,076	0,812 ± 0,049
123	86	0,874 ± 0,036	0,871 ± 0,052	0,883 ± 0,053	0,876 ± 0,036
127	4	0,811 ± 0,045	0,815 ± 0,053	0,809 ± 0,064	0,810 ± 0,046
136	6	0,816 ± 0,036	0,789 ± 0,045	0,868 ± 0,048	0,825 ± 0,033
147	85	0,916 ± 0,034	0,880 ± 0,042	0,966 ± 0,034	0,920 ± 0,032
170	19	0,963 ± 0,017	0,950 ± 0,028	0,979 ± 0,021	0,964 ± 0,017
179	15	0,916 ± 0,024	0,900 ± 0,037	0,938 ± 0,032	0,918 ± 0,023
196	70	0,921 ± 0,023	0,929 ± 0,035	0,913 ± 0,033	0,920 ± 0,023
212	90	0,850 ± 0,026	0,801 ± 0,033	0,933 ± 0,030	0,861 ± 0,023
248	84	0,950 ± 0,016	0,955 ± 0,023	0,947 ± 0,026	0,950 ± 0,016
280	46	0,981 ± 0,012	0,979 ± 0,014	0,983 ± 0,018	0,981 ± 0,012
292	92	0,963 ± 0,014	0,943 ± 0,023	0,987 ± 0,014	0,964 ± 0,013
300	7	0,961 ± 0,015	0,939 ± 0,025	0,987 ± 0,014	0,962 ± 0,015
300	18	0,952 ± 0,017	0,928 ± 0,026	0,980 ± 0,015	0,953 ± 0,016
300	32	0,937 ± 0,018	0,918 ± 0,025	0,960 ± 0,019	0,938 ± 0,017
300	41	0,936 ± 0,017	0,907 ± 0,027	0,972 ± 0,019	0,938 ± 0,016
300	51	0,959 ± 0,014	0,940 ± 0,024	0,983 ± 0,014	0,960 ± 0,013
300	71	0,962 ± 0,013	0,940 ± 0,021	0,988 ± 0,010	0,963 ± 0,012
300	74	0,978 ± 0,012	0,965 ± 0,020	0,992 ± 0,009	0,978 ± 0,012
300	89	0,953 ± 0,014	0,950 ± 0,021	0,957 ± 0,019	0,953 ± 0,014
300	96	0,991 ± 0,008	0,991 ± 0,010	0,991 ± 0,013	0,991 ± 0,008
300	103	0,914 ± 0,020	0,880 ± 0,028	0,959 ± 0,021	0,918 ± 0,019

Tabela 7.12 - *AdaBoost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
12	78	0,501 ± 0,013	0,010 ± 0,100	0,003 ± 0,025	0,004 ± 0,040
13	76	0,519 ± 0,051	0,150 ± 0,359	0,040 ± 0,099	0,063 ± 0,152
16	64	0,608 ± 0,132	0,490 ± 0,502	0,216 ± 0,263	0,287 ± 0,323
18	45	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
25	107	0,729 ± 0,095	1,000 ± 0,000	0,459 ± 0,191	0,605 ± 0,183
36	35	0,690 ± 0,062	0,990 ± 0,100	0,379 ± 0,123	0,538 ± 0,138
39	47	0,816 ± 0,083	1,000 ± 0,000	0,632 ± 0,167	0,761 ± 0,132
43	61	0,678 ± 0,067	1,000 ± 0,000	0,357 ± 0,134	0,512 ± 0,149
53	59	0,833 ± 0,060	1,000 ± 0,000	0,666 ± 0,119	0,793 ± 0,091
82	36	0,895 ± 0,041	1,000 ± 0,000	0,790 ± 0,082	0,880 ± 0,053
89	37	0,811 ± 0,043	1,000 ± 0,000	0,622 ± 0,085	0,764 ± 0,065
103	27	0,937 ± 0,056	0,999 ± 0,007	0,874 ± 0,111	0,928 ± 0,069
117	87	0,566 ± 0,027	0,988 ± 0,103	0,133 ± 0,053	0,230 ± 0,082
123	86	0,643 ± 0,033	0,998 ± 0,013	0,286 ± 0,067	0,441 ± 0,081
127	4	0,635 ± 0,036	1,000 ± 0,000	0,271 ± 0,072	0,421 ± 0,091
136	6	0,761 ± 0,031	1,000 ± 0,005	0,523 ± 0,062	0,685 ± 0,054
143	25	0,907 ± 0,025	1,000 ± 0,003	0,813 ± 0,049	0,896 ± 0,030
170	19	0,929 ± 0,025	0,999 ± 0,005	0,859 ± 0,050	0,923 ± 0,029
179	15	0,800 ± 0,030	0,997 ± 0,008	0,602 ± 0,061	0,749 ± 0,047
196	70	0,872 ± 0,024	0,999 ± 0,004	0,744 ± 0,047	0,852 ± 0,031
212	90	0,744 ± 0,031	0,997 ± 0,009	0,489 ± 0,061	0,654 ± 0,055
248	84	0,953 ± 0,016	1,000 ± 0,001	0,907 ± 0,032	0,951 ± 0,018
280	46	0,746 ± 0,247	0,500 ± 0,502	0,492 ± 0,495	0,496 ± 0,499
292	92	0,978 ± 0,014	1,000 ± 0,002	0,955 ± 0,027	0,977 ± 0,014
300	7	0,988 ± 0,010	1,000 ± 0,000	0,977 ± 0,020	0,988 ± 0,010
300	18	0,894 ± 0,021	0,998 ± 0,006	0,789 ± 0,042	0,881 ± 0,027
300	26	0,931 ± 0,021	0,999 ± 0,004	0,863 ± 0,041	0,926 ± 0,024
300	32	0,881 ± 0,021	0,999 ± 0,004	0,762 ± 0,043	0,864 ± 0,028
300	41	0,987 ± 0,010	1,000 ± 0,001	0,975 ± 0,020	0,987 ± 0,011
300	51	0,974 ± 0,011	1,000 ± 0,002	0,948 ± 0,023	0,973 ± 0,012
300	71	0,944 ± 0,017	0,999 ± 0,004	0,889 ± 0,033	0,940 ± 0,019
300	89	0,957 ± 0,015	1,000 ± 0,001	0,914 ± 0,030	0,955 ± 0,016
300	96	0,992 ± 0,007	1,000 ± 0,001	0,984 ± 0,014	0,992 ± 0,007
300	103	0,848 ± 0,025	0,997 ± 0,007	0,699 ± 0,050	0,821 ± 0,035

Tabela 7.13 - *Robust Boost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,560 ± 0,163	0,120 ± 0,327	0,120 ± 0,327	0,120 ± 0,327
5	67	0,555 ± 0,157	0,110 ± 0,314	0,110 ± 0,314	0,110 ± 0,314
11	52	0,503 ± 0,023	0,020 ± 0,141	0,007 ± 0,047	0,010 ± 0,070
12	78	0,763 ± 0,122	0,940 ± 0,239	0,525 ± 0,245	0,651 ± 0,240
13	76	0,541 ± 0,071	0,290 ± 0,456	0,083 ± 0,142	0,126 ± 0,206
16	64	0,753 ± 0,131	0,970 ± 0,171	0,506 ± 0,261	0,631 ± 0,241
25	107	0,694 ± 0,102	0,980 ± 0,141	0,387 ± 0,203	0,528 ± 0,210
36	35	0,670 ± 0,066	0,980 ± 0,141	0,340 ± 0,133	0,492 ± 0,154
39	47	0,792 ± 0,090	1,000 ± 0,000	0,584 ± 0,180	0,721 ± 0,148
43	61	0,668 ± 0,064	1,000 ± 0,000	0,335 ± 0,128	0,488 ± 0,150
53	59	0,811 ± 0,062	1,000 ± 0,000	0,623 ± 0,124	0,760 ± 0,097
82	36	0,876 ± 0,039	1,000 ± 0,000	0,752 ± 0,079	0,856 ± 0,052
89	37	0,781 ± 0,056	0,999 ± 0,006	0,563 ± 0,111	0,713 ± 0,100
103	27	0,955 ± 0,028	1,000 ± 0,003	0,911 ± 0,055	0,953 ± 0,031
117	87	0,549 ± 0,028	0,928 ± 0,257	0,097 ± 0,055	0,173 ± 0,091
123	86	0,626 ± 0,055	0,975 ± 0,142	0,254 ± 0,110	0,391 ± 0,147
127	4	0,621 ± 0,043	0,987 ± 0,101	0,244 ± 0,088	0,383 ± 0,120
136	6	0,735 ± 0,049	0,998 ± 0,020	0,471 ± 0,096	0,634 ± 0,097
147	85	0,794 ± 0,051	0,999 ± 0,004	0,588 ± 0,103	0,735 ± 0,084
170	19	0,917 ± 0,029	1,000 ± 0,003	0,834 ± 0,057	0,908 ± 0,034
179	15	0,785 ± 0,033	0,995 ± 0,013	0,573 ± 0,066	0,725 ± 0,053
196	70	0,852 ± 0,032	0,998 ± 0,007	0,705 ± 0,063	0,825 ± 0,044
212	90	0,708 ± 0,054	0,985 ± 0,100	0,418 ± 0,108	0,579 ± 0,119
248	84	0,950 ± 0,017	1,000 ± 0,002	0,901 ± 0,035	0,947 ± 0,019
280	46	0,986 ± 0,011	1,000 ± 0,002	0,973 ± 0,022	0,986 ± 0,011
292	92	0,973 ± 0,013	0,999 ± 0,003	0,946 ± 0,027	0,972 ± 0,014
300	7	0,982 ± 0,011	1,000 ± 0,000	0,964 ± 0,021	0,982 ± 0,011
300	18	0,848 ± 0,042	0,997 ± 0,007	0,698 ± 0,083	0,818 ± 0,063
300	26	0,917 ± 0,033	0,999 ± 0,004	0,834 ± 0,066	0,907 ± 0,046
300	32	0,860 ± 0,025	0,997 ± 0,007	0,721 ± 0,049	0,836 ± 0,033
300	41	0,980 ± 0,013	1,000 ± 0,001	0,960 ± 0,025	0,979 ± 0,013
300	51	0,965 ± 0,014	0,999 ± 0,003	0,930 ± 0,028	0,963 ± 0,015
300	71	0,934 ± 0,020	0,998 ± 0,004	0,870 ± 0,039	0,929 ± 0,023
300	89	0,947 ± 0,023	0,999 ± 0,002	0,895 ± 0,047	0,944 ± 0,028
300	96	0,988 ± 0,009	1,000 ± 0,000	0,976 ± 0,017	0,988 ± 0,009
300	103	0,730 ± 0,144	0,846 ± 0,357	0,462 ± 0,289	0,563 ± 0,346

Tabela 7.14 - NNS - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando 36 usuários (um terço do total) ordenados pelas quantidades de ciclos reconhecidos.

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,645 ± 0,228	0,290 ± 0,456	0,290 ± 0,456	0,290 ± 0,456
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,557 ± 0,119	0,225 ± 0,417	0,117 ± 0,239	0,148 ± 0,286
12	78	0,708 ± 0,172	0,720 ± 0,451	0,415 ± 0,345	0,501 ± 0,363
16	64	0,664 ± 0,237	0,330 ± 0,473	0,330 ± 0,473	0,330 ± 0,473
25	107	0,771 ± 0,180	0,849 ± 0,359	0,544 ± 0,358	0,625 ± 0,351
36	35	0,561 ± 0,087	0,458 ± 0,493	0,127 ± 0,174	0,188 ± 0,240
39	47	0,929 ± 0,090	0,999 ± 0,009	0,858 ± 0,180	0,911 ± 0,125
43	61	0,803 ± 0,103	0,989 ± 0,041	0,612 ± 0,201	0,736 ± 0,173
53	59	0,515 ± 0,026	0,310 ± 0,465	0,031 ± 0,053	0,055 ± 0,092
82	36	0,785 ± 0,078	1,000 ± 0,000	0,570 ± 0,155	0,714 ± 0,132
89	37	0,520 ± 0,024	0,520 ± 0,502	0,039 ± 0,049	0,072 ± 0,086
102	83	0,577 ± 0,048	0,970 ± 0,171	0,155 ± 0,095	0,256 ± 0,138
103	27	0,955 ± 0,043	0,999 ± 0,006	0,911 ± 0,085	0,951 ± 0,049
117	87	0,544 ± 0,034	0,892 ± 0,266	0,092 ± 0,066	0,161 ± 0,107
123	86	0,645 ± 0,054	0,997 ± 0,013	0,291 ± 0,109	0,440 ± 0,133
136	6	0,656 ± 0,054	0,989 ± 0,035	0,316 ± 0,108	0,469 ± 0,124
147	85	0,800 ± 0,059	0,994 ± 0,014	0,603 ± 0,118	0,744 ± 0,092
170	19	0,970 ± 0,024	0,996 ± 0,009	0,943 ± 0,049	0,968 ± 0,026
179	15	0,651 ± 0,041	0,995 ± 0,016	0,304 ± 0,083	0,460 ± 0,097
196	70	0,633 ± 0,034	0,998 ± 0,011	0,267 ± 0,069	0,416 ± 0,086
212	90	0,721 ± 0,047	0,980 ± 0,028	0,452 ± 0,094	0,613 ± 0,092
248	84	0,631 ± 0,038	0,999 ± 0,007	0,261 ± 0,075	0,409 ± 0,095
270	9	0,605 ± 0,027	1,000 ± 0,000	0,211 ± 0,055	0,345 ± 0,076
292	92	0,911 ± 0,034	0,991 ± 0,012	0,830 ± 0,067	0,902 ± 0,041
300	7	0,938 ± 0,025	0,996 ± 0,007	0,880 ± 0,049	0,934 ± 0,029
300	18	0,722 ± 0,039	0,987 ± 0,020	0,449 ± 0,077	0,614 ± 0,075
300	26	0,863 ± 0,034	0,995 ± 0,008	0,728 ± 0,069	0,839 ± 0,047
300	32	0,767 ± 0,038	0,994 ± 0,010	0,538 ± 0,076	0,695 ± 0,066
300	41	0,941 ± 0,025	0,998 ± 0,005	0,883 ± 0,049	0,936 ± 0,029
300	51	0,953 ± 0,021	0,993 ± 0,009	0,912 ± 0,043	0,950 ± 0,024
300	71	0,973 ± 0,018	0,997 ± 0,006	0,948 ± 0,036	0,972 ± 0,019
300	74	0,783 ± 0,039	0,998 ± 0,007	0,567 ± 0,079	0,720 ± 0,065
300	89	0,677 ± 0,039	1,000 ± 0,000	0,353 ± 0,079	0,517 ± 0,085
300	96	0,948 ± 0,030	1,000 ± 0,000	0,896 ± 0,060	0,944 ± 0,034
300	103	0,808 ± 0,037	0,998 ± 0,006	0,617 ± 0,073	0,760 ± 0,056

Algumas percepções a serem indicadas sobre os valores das métricas dos usuários com base nas Tabelas 7.11, 7.12, 7.13 e 7.14 e no Apêndice A, por classificador:

- Em relação ao classificador SVM:
    - Os usuários de uma primeira faixa de valores de ciclos reconhecidos ($NC_u < 16$) apresentam, em linhas gerais, valores ruins (várias $Metr < 0{,}6$);
    - em uma segunda faixa, com valores de ciclos reconhecidos conforme $16 \leq NC_u \leq 42$, há uma grande melhoria nos valores de todas as métricas, entretanto ainda são apresentados alguns valores inferiores a 0,8;
    - para $42 < NC_u \leq 300$, praticamente todas as métricas apresentam valores superiores a 0,8.
- No tocante ao classificador *Ada*:
    - Os usuários de uma primeira faixa de valores de ciclos reconhecidos ($NC_u < 11$) apresentam valores muito ruins ($Acu = 0{,}5$ e demais $Metr = 0$);
    - em uma segunda faixa, com valores de ciclos reconhecidos conforme $11 \leq NC_u \leq 18$, há uma leve melhoria nos valores de todas as métricas, entretanto os valores ainda são considerados ruins (várias $Metr < 0{,}6$);
    - para $18 < NC_u \leq 127$, há uma melhoria considerável de *Acu*, *Sen* e *F1s* (métricas variam entre 0,6 e 0,8) e um aumento grande de *Pre* (muitos valores chegam, inclusive, a 1);
    - para $127 < NC_u \leq 300$, praticamente todas as métricas apresentam valores superiores a 0,8 (com algumas exceções que regridem a valores inferiores a 0,5) e, em especial, *Pre* continua apresentando resultados altos (vários valores iguais ou próximos a 1).
- Sobre o classificador *Rob*:
    - Os usuários de uma primeira faixa de valores de ciclos considerada ($NC_u < 13$) apresentam valores muito ruins (quase todas $Acu < 0{,}6$ e demais $Metr$ com vários valores inferiores a 0,1);
    - em uma segunda faixa, com valores de ciclos reconhecidos conforme $13 \leq NC_u \leq 36$, há uma leve melhoria nos valores de *Acu*, *Sen* e *F1s*, e uma melhoria muito grande de *Pre* que apresenta vários valores próximos

ou iguais a 1;

- para $36 < NC_u \leq 127$, há uma melhoria considerável de *Acu*, *Sen* e *F1s* (métricas variam entre 0,6 e 0,8) e *Pre* continua apresentando um resultado excelente (vários valores próximos ou iguais a 1);
- para $127 < NC_u \leq 300$, *Acu*, *Sen* e *F1s* apresentam praticamente todos os valores superiores a 0,8 e *Pre* mantém os seus excelentes resultados.

- Em relação ao classificador NNS:
  - Os usuários de uma primeira faixa de valores de ciclos reconhecidos ($NC_u < 16$) apresentam quase todos os valores muito ruins ($Acu = 0{,}5$ e demais $Metr = 0$);
  - na segunda faixa ($16 \leq NC_u \leq 102$), há uma melhoria nos valores de todas as métricas (vários valores superiores a 0,7), entretanto ainda há uma ocorrência muito grande de valores muito ruins (menores que 0,5) em *Sen* e *F1s*;
  - para $102 < NC_u \leq 270$, há uma pequena melhoria de *Acu*, *Sen* e *F1s* (quase todos os valores de *Acu* variam entre 0,55 e 0,8, e os valores de *Sen* e *F1s* ainda apresentam muitos valores menores que 0,5) e uma melhoria acentuada de *Pre* (vários valores próximos ou iguais a 1);
  - para $270 < NC_u \leq 300$, *Acu*, *Sen* e *F1s* melhoram e apresentam vários valores superiores a 0,8, e *Pre* mantém os seus excelentes resultados.

Os valores indicados nas Tabelas 7.9 e 7.10 agregados às diversas percepções mencionadas apontam para a conclusão de que o classificador SVM apresenta resultados mais adequados que os demais para esse problema de classificação.

Especial atenção deve ser dada, entretanto, aos vários valores bastante altos (próximos ou iguais a 1) que *Pre* apresentou nos outros 3 classificadores (vide Figuras A.2, A.3 e A.4 do Apêndice A). Analisando os valores brutos obtidos do processo de subamostragem aleatória, percebe-se que a explicação para esse fenômeno deriva da grande ocorrência, ao serem utilizados esses 3 classificadores, de pequenas quantidades (muitas iguais a zero) de falsos positivos.

Lembrando que o cálculo de *Pre* é dado por (3.2) e, Caso *fp* tenda a zero, *Pre* tenderá a 1.

Apesar dos altos valores de *Pre* indicarem algo favorável para os classificadores em um contexto comparativo, eles não podem ser analisados isoladamente pois nada indicam em relação às quantidades de *vn* e *fn*. Para o caso em análise, observou-se que *fn* apresentou valores altos, o que é desfavorável e não é percebido com a análise exclusiva dos valores de *Pre*. Assim, o mais adequado é observar a combinação de *Acu*, *Pre* e *Sen*.

Análise das quantidades de ciclos a serem considerados em um treinamento

Em relação às quantidades de ciclos necessários para que um classificador seja considerado bem treinado, observa-se, de maneira geral, que todos os classificadores avaliados apresentam valores de métricas melhores à medida em que o número de ciclos empregados nos seus treinamentos aumenta (algo facilmente perceptível nas tabelas e gráficos do Apêndice A, com algumas poucas exceções).

A conclusão indicada anteriormente corrobora a ideia geral existente na área de aprendizado de máquina que aponta para uma proporcionalidade entre a qualidade de um treinamento e o tamanho da amostra utilizada neste treinamento.

No contexto deste trabalho, considerando os limites impostos pela base de dados escolhida, não foi possível verificar se valores superiores a 300 ciclos geram melhores resultados de métricas que os apontados para $NC_u = 300$. De qualquer forma, os resultados registrados na Tabela 7.10, particularmente para a SVM, podem ser considerados muito bons. Conclui-se, então, que é recomendável que sejam utilizados, pelos menos, 300 ciclos para o treinamento do classificador escolhido.

Análise dos custos (tempo) dos classificadores

Durante o processo de subamostragem aleatória, para todos as atividades de treinamento dos classificadores e classificação das amostras de ciclos reconhecidos, foi feita a medição do tempo gasto pelos diversos algoritmos empregados. Essa medição foi realizada com o propósito de viabilizar uma comparação entre os 4 classificadores e verificar qual apresenta melhor desempenho nesse contexto.

As Tabelas A.5, A.5, A.7 e A.8 constantes do Apêndice A indicam as médias dos tempos de treinamento ($\boldsymbol{\mu}(\Delta \boldsymbol{t_{Tr}})$) e de classificação ($\boldsymbol{\mu}(\Delta \boldsymbol{t_{Cl}})$), por usuário e quantidades de ciclos reconhecidos (forma de ordenação).

Analisando essas tabelas, percebe-se que não há uma relação de proporcionalidade entre os tempos de treinamento ($\boldsymbol{\mu(\Delta t_{Tr})}$) e os números de ciclos utilizados $\boldsymbol{NC_u}$. Também não foi observada, em todos os classificadores, qualquer relação entre os valores das métricas e os tempos de treinamento e classificação.

A partir da conclusão apontada na análise das quantidades de ciclos feita anteriormente (utilizar $NC_u \geq 300$), a Figura 7.6 indica uma comparação entre as médias dos tempos de treinamento e classificação, em segundos, considerando cada classificador e todos os usuários cujos $NC_u = 300$.

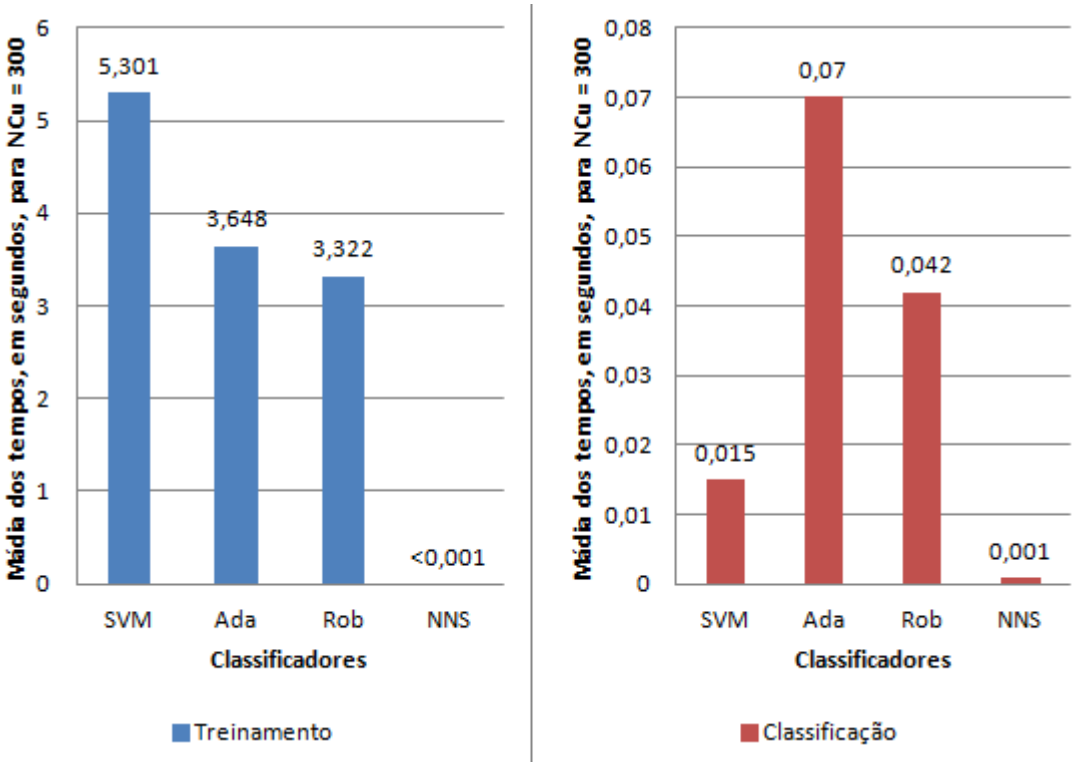

Figura 7.6 - Tempos médios gastos pelas várias amostragens aleatórias para o treinamento e para a classificação

A Figura 7.6 indica que todos os tempos gastos nos processos de classificação são menores que 70 milésimos de segundo. Para o modelo de autenticação proposto neste trabalho, este tempo é considerado desprezível. Para o treinamento dos classificadores, o gráfico indica que o mais rápido é o NNS e o mais lento é o SVM. Apesar de ser o que consome mais tempo para o seu treinamento, o classificador SVM gasta, em média, para o caso analisado, pouco mais de 5 segundos, o que também é considerado, para o contexto

geral dessa tese, bastante adequado para essa atividade (lembrando que o treinamento é uma atividade realizada durante a fase de registro explicada no item "6.3-Definição do Modelo de Autenticação" e que não demanda rapidez extrema).

7.1.4- Análise estatística dos resultados em termos de eficácia do processo

Com o propósito de avaliar a significância estatística dos resultados obtidos nas análises apresentadas nos itens anteriores, são realizados testes de hipóteses considerando os valores de cada uma das métricas para cada um dos classificadores utilizados neste trabalho.

Antes dos testes de hipóteses voltados para a comparação de métricas idênticas entre classificadores diferentes, são feitos testes de normalidade para cada uma das dezesseis combinações possíveis (quatro métricas em quatro classificadores) nos quais são utilizados os resultados de todos os 108 usuários. Esses testes adotam hipóteses nulas (H_0) nas quais as distribuições dos valores das métricas são assumidas como **normais**. Considera-se que H_0 é rejeitada quando os valores de p são baixos, e adotou-se neste trabalho o limite de 0,05, correspondendo ao intervalo de 95 % de confiança.

Os resultados desses testes de normalidade são indicados na Tabela 7.15.

Tabela 7.15 – Valores de *p* correspondentes aos testes de normalidades de cada métrica em cada classificador, considerando todos os 108 usuários.

Métrica / Classificador	**Acu**	**Pre**	**Sen**	**F1s**
SVM	0,04256	0,13259	0,40852	0,10655
Ada	0,00370	0,00100	0,00396	0,00100
Rob	0,00100	0,00100	0,00100	0,00100
NNS	0,01224	0,00100	0,01090	0,00330

De acordo com os valores indicados em vermelho (menores que 0,05) na Tabela 7.15, percebe-se que a maioria das distribuições têm suas normalidades rejeitadas, exceto as da precisão, da sensibilidade e de *F1-score* do classificador SVM.

A partir da aceitação ou da rejeição da normalidade das distribuições, adotam-se testes diferentes para a comparação entre as métricas dos classificadores. Caso ambas as

métricas comparadas não tenham a normalidade rejeitada ($p \geq 5\%$), adota-se um teste paramétrico ("*T-Student*"); por outro lado, caso pelo menos uma das métricas comparadas tenha a normalidade rejeitada, é adotado um teste não paramétrico (teste de "*Wilcoxon*"). Essa é uma escolha típica desses testes, como pode ser observado em [KOMATSU, 2017] e [BARROS, 2005].

Para a comparação entre os classificadores são realizados novos testes de hipóteses nos quais as H_0 representam condições nas quais as métricas medidas derivam de distribuições com a mesma mediana. Para que H_0 seja rejeitada, assume-se novamente que a probabilidade de significância (valor de p) deva ser menor a 5%.

A Tabela 7.16 apresenta os valores de p obtidos na comparação entre as métricas dos classificadores empregados nesta tese.

Tabela 7.16 - Valores de p dos testes de comparação entre os classificadores.

		Ada	***Rob***	***NNS***
Acu	***SVM***	0.06717	0.00477	0,01279
	Ada	-	0.00022	0.00010
	Rob	-	-	0,00009
	NNS	-	-	-
Pre	***SVM***	0.00023	0,02417	0,00113
	Ada	-	0.05390	0.00266
	Rob	-	-	0.19125
	NNS	-	-	-
Sen	***SVM***	0,00123	0,00128	0,00114
	Ada	-	0.00022	0.00013
	Rob	-	-	0.00013
	NNS	-	-	-
F1s	***SVM***	0.01043	0.00070	0,00640
	Ada	-	0.00022	0.00042
	Rob	-	-	0.00039
	NNS	-	-	-

Como é possível perceber, dos vinte e quatro valores de *p* (quatro métricas e seis valores de *p* por métrica), somente três são superiores a 5% (dados em azul na tabela). Essas comparações ("*SVM* versus *Ada"*, "*Ada* versus *Rob"* e "*Rob* versus *NNS"*) são consideradas, dessa forma, as únicas cujos resultados não são estatisticamente significativos. As outras vinte e uma, então, podem ser consideradas comparações estatisticamente significativas.

Esses resultados sugerem que, com respeito à maioria dos casos considerados, há uma diferença estatisticamente significativa entre os desempenhos dos classificadores. A SVM resultou em acurácia estatisticamente superior ao classificador NNS e ao *Robust Boost*. Sua precisão, sensibilidade e *F1-score* foram estatisticamente superiores às correspondentes métricas do *AdaBoost*, à do *Robust Boost* e à do NNS.

7.2- CONCLUSÃO

O presente capítulo tem o objetivo de apresentar uma avaliação da proposta de protocolo de autenticação descrita detalhadamente no capítulo anterior. Essa avaliação focaliza os aspectos que envolvem a eletrocardiografia e o aprendizado de máquina.

São apresentadas informações sobre a base de dados de ECG utilizada e a preparação das *features* utilizadas no modelo de autenticação. Após isso, é detalhado o processo de subamostragem aleatória que fundamenta os cálculos das métricas de análise empregadas na avaliação da eficácia, e dos intervalos de tempo aplicados na avaliação da eficiência das técnicas escolhidas. São apresentadas várias fórmulas, tabelas e gráficos que têm o propósito de deixar clara a metodologia empregada no processo de avaliação da proposta. Por fim, é realizada uma análise estatística baseada em probabilidade de significância.

Como conclusão obtida a partir da análise de eficácia feita com o uso das diversas métricas de análise adotadas, é possível afirmar que, dentre as técnicas de ML escolhidas, a SVM é a que apresenta desempenho mais robusto. Além disso, a análise temporal indica que os custos do protocolo, em termos de tempo, são adequados para o cenário de emprego do protocolo proposto.

8- CONCLUSÕES E RECOMENDAÇÕES

O presente trabalho tem, como objetivo principal, a proposta de um novo protocolo de autenticação forte, mútua e contínua para ambientes de MCC. Esse protocolo usa a eletrocardiografia como a sua fonte de dados biométricos que agregam as características da transparência para o usuário e da continuidade do processo de autenticação.

Além do objetivo principal, também são objetivos desta tese: a realização de revisão bibliográfica sobre vários temas vinculados ao foco do trabalho (computação em nuvem, computação móvel em nuvem, segurança da informação, fundamentos de autenticação, biometria, eletrocardiografia e aprendizado de máquina, dentre outros), a validação e avaliação do desempenho de partes da proposta, e a análise comparativa desta em relação a alguns trabalhos similares.

Voltado para esses objetivos, nesta tese é apresentada uma revisão bibliográfica geral sobre: computação em nuvem, computação móvel em nuvem, fundamentos de segurança da informação e suas aplicações em ambientes de CC, fundamentos da autenticação, técnicas de biometria e sua aplicação para a autenticação, conceitos de autenticação contínua e multimodal, aspectos do funcionamento do coração e da captura de sinais elétricos oriundos deste órgão, fundamentos da biometria baseada em sinais de eletrocardiografia e conceitos de aprendizado de máquina.

Além da revisão bibliográfica geral, são descritos e comparados, à proposta principal deste trabalho, alguns trabalhos relacionados aos temas explorados. Em seguida, é apresentada a proposta de protocolo de autenticação, detalhando as suas diversas fases, os métodos utilizados e alguns modelos empregados.

Por fim, a tese apresenta a avaliação da proposta de protocolo focalizando, particularmente, os aspectos vinculados à eletrocardiografia.

Após todas as pesquisas, discussões, análises e elucubrações, é possível chegar-se às conclusões gerais descritas no item a seguir.

8.1- CONCLUSÕES GERAIS

Acerca das questões de pesquisa indicadas na introdução deste trabalho (e repetidas nesta conclusão), as seguintes considerações devem ser feitas:

- O emprego da eletrocardiografia permite autenticação de forma eficaz e eficiente?

  Sim. No capítulo 4 desta tese, é mostrado que as *features* extraídas de um ECG podem ser consideradas parâmetros biométricos adequados para processos de autenticação de seres humanos. Além disso, a própria proposta de protocolo que é o cerne deste trabalho, aliada a algumas referências bibliográficas consultadas e indicadas no capítulo 5, sugerem modelos a partir dos quais essa autenticação biométrica pode ser realizada de forma produtiva. Há propostas que estão baseadas na interpretação de pontos fiduciais e outras que utilizam domínios transformados obtidos a partir de análises dos ciclos cardíacos registrados nos ECG. Também se observa o emprego de técnicas de reconhecimento de padrões voltadas para a obtenção da eficácia e da eficiência necessárias para um processo de autenticação robusto.

- É possível utilizar alguma técnica de aprendizado de máquina como parte de um processo de autenticação baseado em eletrocardiografia?

  Sim. No capítulo 5 deste trabalho são indicados alguns artigos científicos nos quais há relatos da utilização de técnicas de ML como abordagens consistentes para a resolução de problemas de autenticação. São sugeridos, por exemplo, o uso de SVM e NNS como métodos a serem empregados para, a partir do aprendizado realizado por esses classificadores, a classificação de ciclos cardíacos obtidos de ECG de pessoas com o propósito de autenticá-las. Além dessas propostas encontradas na literatura, o presente trabalho também propõe e avalia, nos capítulo 6 e 7, a utilização das técnicas de ML *AdaBoost* e *Robust Boost* como opções válidas para a resolução do problema abordado.

- Caso seja, quão eficazes são essas técnicas de ML?

  Na análise de resultados tratada no capítulo 7, as métricas de análise referentes à acurácia, precisão, sensibilidade e *F1-score* permitem a avaliação da eficácia das técnicas aplicadas e representam, por exemplo,

níveis de acurácia próximos a 95%, de precisão superiores a 98%, de sensibilidade próximos a 85% (a SVM, nesse contexto, apresenta valor bastante superior às outras técnicas, atingindo 97%) e de *F1-score* da ordem de 90% (a SVM, também em relação à essa métrica, apresenta valor bastante superior às outras técnicas, atingindo 95%). Esses números indicam que, em diversos contextos, as técnicas de ML mostram-se bastante eficazes na resolução do problema em questão.

Além disso, no mesmo capítulo, é feita uma análise dos tempos gastos durante as tarefas de ML que aponta para valores muito baixos, o que indica, pelo menos nesse contexto, uma eficiência adequada.

- É possível projetar um protocolo de autenticação forte, multimodal, contínua e não intrusiva que se baseie em eletrocardiografia?

  Sim. O capítulo 6 apresenta uma proposta de protocolo na qual são utilizados três fatores de autenticação independentes ("algo que se tem, algo que se sabe e algo intrínseco à pessoa"), o que caracteriza uma autenticação forte. Um desses fatores, baseado em biometria fisiológica, permite que um indivíduo permaneça sendo autenticado continuamente e de forma não intrusiva. O tipo de biometria adotado é a eletrocardiografia, que viabiliza sinais cardíacos capturados por sensores justapostos ao corpo do usuário e que não o inabilitam a realizar as suas atividades costumeiras (método não intrusivo). Além disso, a captura e a análise dos sinais cardíacos é realizada de maneira contínua, o que permite a manutenção ininterrupta do relacionamento de confiança existente entre o usuário e o provedor perante o qual o mesmo encontra-se autenticado.

- De que forma a área de segurança da informação aplicada à computação móvel em nuvem pode ser beneficiada pelo emprego da eletrocardiografia?

  Um dos principais problemas de segurança da informação aos quais as áreas de computação em nuvem e computação móvel em nuvem estão submetidas é o ataque de personificação de usuários. Por meio deste ataque, pessoas não autorizadas buscam acesso a recursos e serviços fornecidos por CSP utilizando, para isso, a identidade de pessoas autorizadas. No caso específico de MCC, são utilizados equipamentos móveis a partir dos quais os serviços são acessados mediante algum tipo de autenticação. É típico, por

exemplo, autenticar-se perante um provedor de serviços, nesse contexto, utilizando apenas o IMSI (o que gera, na maior parte das vezes, uma solicitação de confirmação enviada para o UE no qual esse IMSI se encontra). Essa forma de autenticação permite que pessoas acessem os serviços simplesmente pelo fato de terem um UE com o IMSI considerado, não havendo qualquer garantia de que o usuário realmente seja o "dono" daquele IMSI.

No exemplo ilustrado anteriormente, a utilização da eletrocardiografia para a autenticação de usuários, conforme abordado nas outras questões de pesquisa, se contrapõe, de forma inequívoca, à personificação e faz com que esse tipo de técnica biométrica possa ser considerado algo benéfico para o cenário indicado.

Além das conclusões referentes às indagações que existiam no início dessa tese, também é possível concluir que o problema central abordado aqui, qual seja, "Autenticar, eficiente e eficazmente, usuários e provedores de serviços de MCC, uns perante os outros, de maneira pouco intrusiva e contínua, utilizando vários métodos de maneira conjugada para garantir uma autenticação forte", pode ser resolvido com a utilização do protocolo proposto.

Apesar das várias conclusões positivas, faz-se necessário relatar algumas constatações que não são tão favoráveis:

- A utilização, para um esquema de autenticação baseado em eletrocardiografia, de um processo de detecção de pontos fiduciais que apresenta problemas durante o reconhecimento desses pontos, é algo que pode dificultar e até inviabilizar o emprego prático dessa técnica de biometria sobre alguns usuários. Isso pode ser percebido por meio das análises (constantes do capítulo 7) dos resultados baseados em números de ciclos reconhecidos (lembrando que o reconhecimento dos ciclos cardíacos, no contexto deste trabalho, depende diretamente da detecção de pontos fiduciais).
- O fato da base de dados utilizada possuir apenas ECG de 5 minutos não permitiu que se chegasse a uma conclusão precisa em relação à quantidade mínima de ciclos que devem ser reconhecidos para que o processo tenha uma eficácia estabilizada.

- Alguns valores de métricas da proposta desta tese ficaram abaixo de valores análogos apresentados em outras propostas científicas.

Tendo apontado as principais conclusões, faz-se necessário, também, indicar as principais contribuições desta tese que vão além das publicações já indicadas no capítulo de introdução.

8.2- PRINCIPAIS CONTRIBUIÇÕES

Das análises comparativas entre o atual trabalho e outros similares, pode-se constatar que foram feitos avanços em relação às pesquisas e às práticas adotadas sobre algumas das áreas que foram abordadas nesta tese.

De maneira geral, a proposta de um protocolo de autenticação contínua forte combinando três fatores que não foram combinados em outros trabalhos (o IMSI vinculado ao *chip* do usuário de equipamentos móveis, uma senha e os sinais de eletrocardiografia do usuário) é algo que pode ser considerado uma contribuição para a ciência pois apresenta uma nova opção para atacar o problema central indicado neste trabalho.

Especificamente sobre a utilização de sinais de eletrocardiografia de um usuário para a autenticação contínua deste perante um determinado provedor de serviços, merecem destaque os seguintes aspectos considerados inovadores:

- proposta de processo de atualização dos padrões de ECG armazenados na base de autenticação, ou seja, dos parâmetros dos classificadores (vetores de suporte das SVM, *features* médias das NNS, etc.), derivada da ideia de treinamento de máquina de vida inteira (do inglês, *Lifelong Machine Learning*) cuja aplicação direcionada para o problema em questão não foi encontrada nas pesquisas que foram realizadas até o presente momento;
- implementação de um novo algoritmo de detecção de pontos fiduciais com base em padrão de formato pré definido do ciclo cardíaco que, apesar de não ter apresentado resultados excelentes (houve várias falhas na detecção de pontos fiduciais, como comentado nesta tese), forneceu bons resultados em comparação a processos similares indicados na bibliografia;
- utilização da distância *LP-TP* como parâmetro de normalização, ao invés da

distância *R-R* observada em outros trabalhos, que gerou *features* de boa qualidade para os processos de ML;

- utilização do *AdaBoost* e do *Robust Boost* como opções para a classificação de sinais de eletrocardiografia;
- aplicação da técnica de subamostragem aleatória para a validação do processo de autenticação baseado em eletrocardiografia;
- utilização, vinculada à técnica de subamostragem aleatória, de várias métricas de análise (acurácia, precisão, sensibilidade e *f1-score*) para a avaliação dos processos de classificação empregados sobre o problema central desta tese, ao invés de focalizar apenas a acurácia ou as taxas de acertos e erros. Esse aspecto, apesar de não ser totalmente inovador (foi encontrado em apenas um artigo pesquisado relacionado ao problema), merece ser destacado, pois se mostra como diferencial em relação à quase totalidade dos trabalhos analisados.

8.3- RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Durante as atividades realizadas para a elaboração desta tese, surgiram algumas ideias de temas que podem ser explorados em trabalhos a serem realizados no futuro. Seguem, então, algumas sugestões para esse contexto:

- Empregar dos classificadores utilizados no presente trabalha alterando-se alguns parâmetros utilizados como, por exemplo, o *kernel* da SVM.
- Realizar comparação de resultados obtidos sobre os mesmos ECG e que são oriundos da aplicação de diferentes medidas (RR, LP-TP, etc.) para o processo de normalização das *features*.
- Melhorar a qualidade do algoritmo de detecção de pontos fiduciais de tal maneira que as quantidades de ciclos reconhecidos sejam maiores.
- Elaborar novo algoritmo de detecção de pontos fiduciais que não focalize somente os ECG com ciclos que seguem o modelo indicado neste trabalho.
- Elaborar processo de autenticação baseado em eletrocardiografia que utilize, além das *features* de ECG que não apresentam anomalias, outras características derivadas, inclusive, de problemas cardíacos.
- Simular a proposta de protocolo em um simulador de redes para avaliar o seu funcionamento.

REFERÊNCIAS BIBLIOGRÁFICAS

[ABDULWAHID, 2015] - AL ABDULWAHID, A., CLARKE, N., STENGEL, I., FURNELL, S., REICH, C. (2015). "Continuous and Transparent Multimodal Authentication: Reviewing the State of the Art". Cluster Computing, Vol. 19, No. 1.

[ABO-ALIAN, 2016] - ABO-ALIAN, A., BADR, N. L., TOLBA, M. F. (2016). "Keystroke Dynamics-Based User Authentication Service for Cloud Computing". Concurrency and Computation, Practice and Experience, Vol. 28, No. 9.

[ABRAMS, 1995] - ABRAMS, M. D., JOYCE, M. V. (1995). "Trusted System Concepts". Computers & Security, Vol. 14, No. 1.

[ALBUQUERQUE, 2010] - ALBUQUERQUE, S. L., GONDIM, P. R. L., MONTEIRO, C. C. (2010). "Aspectos de Segurança na Interconexão de Redes Celulares e WLANs". SBSeg2010 - 10o. Simpósio Brasileiro de Segurança da Informação.

[ALBUQUERQUE, 2012] - ALBUQUERQUE, S. L., GONDIM, P. R. L. (2012). "Applying continuous authentication to protect electronic transactions". Capítulo do Livro "Information Assurance and Security Technologies for Risk Assessment and Threat Management: Advances. ", IGI Global.

[ALBUQUERQUE, 2016] - ALBUQUERQUE, S. L., GONDIM, P. R. L. (2016). "Security in Cloud-Computing-Based Mobile Health". IT Professional, Vol. 18, No. 3.

[ALIZADEH, 2016] - ALIZADEH, M., ABOLFAZLI, S., ZAMANI, ., BAAAHARUN, S., SAKURAI, K. (2016). "Authentication in Mobile Cloud Computing: A Survey". Journal of Network and Computer Applications, Vol. 61, No. C.

[AL-RUBAIE, 2016] - AL-RUBAIE, M., CHANG, J. M. (2016). "Reconstruction Attacks Against Mobile-Based Continuous Authentication Systems in the Cloud". Information Forensics and Security, IEEE Transactions On, Vol. 11, No. 12.

[ANAND, 2010] - ANAND, R., BAJPAI, G., BHASKAR, V. (2010). "3D Signature for Efficient Authentication in Multimodal Biometric Security Systems". IACSIT - International Journal of Engineering and Technology, Vol. 2, No. 2.

[AUDITHAN, 2017] - AUDITHAN, S., VIJAYASARO, V., VIJAYAKUMAR, P., VIJAYAKUMAR, V. (2017). "An Efficient Authentication Scheme for Mobile Cloud Computing Services". Journal of Information Science and Engineering, Vol. 33, No. 3.

[AZZINI, 2008] - AZZINI, A., MARRARA, S., SASSI, R., SCOTTI, F. (2008). "A fuzzy approach to multimodal biometric continuous authentication". Fuzzy Optimization and Decision Making , Vol. 7, No. 3.

[BABAEIZADEH, 2014] - BABAEIZADEH, M., BAKHTIARI, M., MAAROF, M. (2014) "Keystroke Dynamic Authentication in Mobile Cloud Computing". International Journal of Computer Applications, Vol. 90, No. 1.

[BAHL, 2012] BAHL, P., HAN, R., Y., LI, L., E., SATYANARAYANAN, M. (2012). "Advancing the State of Mobile Cloud Computing". 3rd ACM workshop on Mobile cloud computing and services, Vol.21, No. 28.

[BARROS, 2005] - BARROS, E., MAZUCHELI, J. (2005). "Um estudo sobre o tamanho e poder dos testes t-Student e Wilcoxon". Acta Scientiarum: Language and Culture, Vol. 27, No. 1.

[BHAVSAR, 2012] - BHAVSAR, H., PANCHAL, M. H. (2012). "A Review on Support Vector Machine for Data Classification". IJARCET - International Journal of Advanced Research in Computer Engineering & Technology , Vol. 1, No. 10.

[BOTTOU, 2016] - BOTTOU, L., CURTIS, F. E., NOCEDAL, J. (2016). "Optimization Methods for Large-Scale Machine Learning". Society for Industrial and Applied Mathematics, Vol. 60, No. 2.

[CALDERON, 2006] - CALDERON, T. G., CHANDRA, A., CHEH, J. J. (2006). "Modeling an intelligent continuous authentication system to protect financial information resources". International Journal of Accounting Information Systems, Vol 7.

[CAMARA, 2017] - CAMARA, C., PERIS-LOPEZ, P., GONZALEZ-MANZANO, L., TAPIADOR, J. (2017). "Real-Time Electrocardiogram Streams for Continuous Authentication". Applied Soft Computing Journal, Vol. 68.

[CHEN, 2017] - CHEN, J., PENG, H., RAZI, A. (2017). "Remote ECG Monitoring Kit to Predict Patient-Specific Heart Abnormalities". Journal of Systemics, Vol. 15, No. 4.

[CLARKE, 2002a] - CLARKE, N., FURNELL, S., RODWELL, P., REYNOLDS, P. (2002). "Acceptance of Subscriber Authentication for Mobile Telephony Devices". Computers & Security, Vol. 21, No. 3.

[CLARKE, 2002b] - CLARKE, N., FURNELL, S., RODWELL, P., REYNOLDS, P. (2002). "Biometric Authentication for Mobile Devices". 3rd Australian Information Warfare and Security Conference.

[CLARKE, 2007] - CLARKE, N., FURNELL, S. (2007). "Advanced user authentication for mobile devices". Computers & Security, Vol. 26.

[CORTES, 1995] - CORTEZ, C., VAPNIK, V. (1995). "Support-Vector Networks". Machine learning, Vol. 20, No. 3.

[DATTA, 2003] - DATTA, A., HAUSWIRTH, M. (2003). "Beyond 'Web of Trust': Enabling P2P E-commerce". ICEC 2003 - IEEE International Conference on Electronic E-commerce.

[DEEPAK, 2015] - DEEPAK, S., GOUTHAM, N. S. (2015). "Face Recognition using Cloud Based Security in Mobile Devices". International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, No. 6.

[DEY, 2016] - DEY, S., SAMPALLI, S., YE, Q. (2016). "MDA: Message Digest-Based Authentication for Mobile Cloud Computing". Journal of Cloud Computing, Vol. 5, No. 1.

[DINH, 2013] - DINH, H. T., LEE, C., NIYATO, D., WANG, P., (2013). "A survey of mobile cloud computing: architecture, applications, and approaches". Wireless Communication Mobile Computing, Vol. 13.

[ETSI TS 133 2014] - "Universal Mobile Telecommunications System (UMTS); LTE; 3G security; Wireless Local Area Network (WLAN) interworking security". V. 12.1.0, Release 12, 2014.

[FALCONI, 2016] - FALCONI, J. S. A., OSMAN, H. A., SADDIK, A. E. (2016). "ECG Authentication for Mobile Devices". Instrumentation and Measurement, IEEE Transactions On, Vol. 65, No. 3.

[FALCONI, 2018] - FALCONI, J. S. A., OSMAN, H. A., SADDIK, A. E. (2018). "ECG and Fingerprint Bimodal Authentication". Sustainable Cities and Society, Vol. 40.

[FAN, 2011] - FAN, X., CAO, J., MAO, H. (2011). "A Survey of Mobile Cloud Computing". ZTE Communications Magazine, No. 1.

[FERRAG, 2019] - FERRAG, M. A., MAGLARAS, L., DERHAB, A. (2019) - "Authentication and Authorization for Mobile IoT Devices Using Biofeatures: Recent Advances and Future Trends". Hindawi Security and Communication Networks, Vol. 2019.

[FURNELL, 2000] - FURNELL, S. M., DOWLAND, P. S., ILLINGWORTH, H., M., REYNOLDS, P., L. (2000). "Authentication and Supervision: a Survey of User Attitudes". Computers & Security, Vol. 19, No. 6.

[FURNELL, 2008] - FURNELL, S. M., CLARKE, N., KARATZOUNI, S. (2008). "Beyond the PIN: Enhancing user authentication for mobile devices". Computer Fraud & Security, No. 8.

[GENTRY, 2009] - GENTRY, C. (2009). "A fully homomorphic encryption scheme", Ph.D. Theses, Stanford University, CA, USA.

[GIURGIU, 2009] - GIURGIU, I., RIVA, O., JURIC, D., KRIVULEV, I., ALONSO, G. (2009). "Calling the cloud: enabling mobile phones as interfaces to cloud applications". ACM International Conference on Distributed Systems Platforms and Open Distributed Processing.

[GORMAN, 2003] - O'GORMAN, L. (2003). "Comparing Passwords, Tokens, and Biometrics for User Authentication". IEEE Proceedings, Vol. 91, No. 12.

[GOUDELIS, 2005] - GOUDELIS, G., TEFAS, A., PITAS, I. (2005). "On Emerging Biometric Technologies". 3rd COST 275 Workshop - Biometrics on the Internet.

[GOYAL, 2014] - GOYAL, S. (2014). "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", International Journal of Computer Network and Information Security, Vol. 6, No. 3.

[GUNSON, 2011] - N. GUNSON, D. MARSHALL, F. MCINNES, M. JACK (2011) "Usability evaluation of voiceprint authentication in automated telephone banking: Sentences versus digits". UK Oxford Journals Science & Mathematics & Social Sciences Interacting with Computers, Vol. 23, No. 1.

[HAMAMI, 2015] - AL-HAMAMI, A. H., AL-JUNEIDI, J. Y. (2015). "Secure Mobile Cloud Computing Based-On Fingerprint". WCSIT - World of Computer Science and Information Technology Journal, Vol. 5, No. 2.

[HAYAJNEH, 2016] - HAYAJNEH, T., MOHD, B. J., IMRAN, M., ALMASHAQBEH, G., VASILAKOS A. V. (2016), "Secure Authentication for Remote Patient Monitoring with Wireless Medical Sensor Networks". Sensors (Basel), Vol. 16, No. 4.

[HE, 2018] - HE, D., KUMAR , N., SHEN, J. (2018). "Efficient Privacy-Aware Authentication Scheme for Mobile Cloud Computing Services". IEEE Systems Journal, Vol. 12, No. 2.

[HELIB, 2014] “Helib An Implementation of homomorphic encryption,” (2014). Website disponível em: https://github.com/shaih/HElib. Acessado em 18 de junho de 2017.

[HRESTAK, 2014] - HRESTAK, D., PICEK, S. (2014). "Homomorphic Encryption in the Cloud", MIPRO 2014 - 37th International Convention on Information and Communication Technology, Electronics and Microelectronics.

[IKEHARA, 2010] - IKEHARA, C. S., CROSBY, M. E. (2010). "Physiological Measures Used for Identification of Cognitive States and Continuous Authentication". CHI2010 - 28th ACM Conference on Human Factors in Computing Systems.

[INDRAWAN, 2013] - INDRAWAN, P., BUDIYATNO, S., RIDHO, N. M., SARI R. F. (2013). "Face Recognition for Social Media With Mobile Cloud Computing". IJCCSA - International Journal on Cloud Computing: Services and Architecture, Vol. 3, No. 1.

[ISLAM, 2017] - ISLAM, S, AMMOUR, N., ALAJLAN, N., WADUD, M. A. (2017). "Selection of Heart-Biometric Templates for Fusion." IEEE Access, Vol. 5.

[JAIN, 2004] - JAIN, A. K., ROSS, A., PRABHAKAR, S. (2004). "An Introduction to Biometric Recognition". IEEE Transactions on Circuits and Systems for Video Technology, Vol.14, No. 1.

[JANNATI, 2017] - JANNATI, B. (2017). "An Improved Authentication Protocol for Distributed Mobile Cloud Computing Services". International Journal of Critical Infrastructure Protection, Vol. 19.

[JEONG, 2014] - JEONG, Y. S., PARK, J. S., PARK, J. H. (2014). "An Efficient Authentication System of Smart Device Using Multi Factors in Mobile Cloud Service Architecture." International Journal of Communication Systems, Vol. 28, No. 4.

[JIANG, 2016] - JIANG, Q., MA, J., WEI, F. (2016), "On the security of a privacy-aware authentication scheme for distributed mobile cloud computing services", IEEE Systems Journal, Vol. PP, No. 99.

[JOSANG, 2007] - JOSANG, A., ISMAIL, R., BOYD, C. (2007). "A survey of trust and reputation systems for online service provision". Decision Support Systems, Vol. 43, No. 2.

[KHAN, 2014] KHAN, A. R., OTHMAN, M., MADANI, S. A., KHAN, S. U. (2014). "A Survey of Mobile Cloud Computing Application Models", IEEE Communications Surveys & Tutorials, Vol. 16, No. 1.

[KLOSTERMAN, 2000] - KLOSTERMAN, A. J., GANGER, G. R. (2000). "Secure Continuous Biometric-Enhanced Authentication". CMU-CS-00-134 - CMU SCS Technical Report.

[KOHAVI, 1995] - KOHAVI, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection". IJCAI - International Joint Conference on Articial Intelligence.

[KOHAVI, 1998] - KOHAVI, R., PROVOST, F. (1998) "Glossary of terms". Machine Learning - Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. Vol. 30.

[KOIEN, 2003] - KOIEN, G. M., HASLESTAD, T. (2003). "Security Aspects of 3G-WLAN Interworking". IEEE Communications Magazine, Vol. 41, No. 11.

[KOMATSU, 2017] - KOMATSU, A. (2017). "Comparação dos poderes dos teste t de Student e Mann-Whitney Wilcoxon pelo método de Monte Carlo". Revista da Estatística UFOP, Vol. VI.

[KOMORI, 2011] - KOMORI, O., EGUCHI, S. (2011). "Boosting Learning Algorithm for Pattern Recognition and Beyond". IEICE Transactions. Vol. 94, No. D.

[KOVACHEV, 2011] - KOVACHEV, D., CAO, Y., KLAMMA, R., (2011). "Mobile Cloud Computing: A Comparison of Application Models". Informatik 5 Information Systems.

[LAGUNA, 1997] - LAGUNA, P., MARK, R. G., GOLDBERG, A., MOODY G. B. (1997). “A database for evaluation of algorithms for measurement of QT and other wave-form intervals in the ECG,” Computing in Cardiology.

[LEE, 2018] - LEE, H., HWANG, J. Y., KIM, D., LEE, S., LEE, S. H., SHIN, J. S. (2018). " Understanding Keystroke Dynamics for Smartphone Users Authentication and Keystroke Dynamics on Smartphones Built-In Motion Sensors". Applied Cryptography and Noise Resistant Data Security, Vol. 2018.

[LIN, 2015] - LIN, H., XU, L., HUANG, X., WU, W., HUANG, Y. (2015). "A Trustworthy Access Control Model for Mobile Cloud Computing Based on Reputation and Mechanism Design". Ad Hoc Networks, Vol. 35, No. C.

[LIN, 2016] - LIN, H., HU, J., TIAN, Y., YANG, L., XU, L. (2016). "Toward Better Data Veracity in Mobile Cloud Computing: A Context-Aware and Incentive-Based Reputation Mechanism". Information Sciences, Vol. 387, No. C.

[LIN, 2017] - LIN, F., SONG, C., ZHUANG, Y., XU, W., LI, C., REN, K. (2017). "Cardiac Scan: A Non-Contact and Continuous Heart-Based User Authentication System," Mobicom - IEEE International Conference on Mobile Computing and Networking.

[LIU, 2006] - LIU, J., JIANG, S., LIN, H. (2006). "Introduction to Diameter – Get the next generation AAA protocol". Website disponível em "http://www.ibm.com/developerworks/wireless/library/wi-diameter". Acessado em 25 de abril de 2017.

[LOUK, 2015] - LOUK, M., LIM, H. (2015). "Homomorphic Encryption in Mobile Multi Cloud Computing". ICOIN - 2015 International Conference on Information Networking .

[PATRO, 2017] - PATRO, K., KUMAR., P. (2017). "Effective Feature Extraction of ECG for Biometric Application". Procedia Computer Science, Vol. 115.

[MASON, 2016] - MASON, J. W., BADILINI, F., VAGLIO, M., LUX, R. L., AYSIN, B., MOON, T. E., HEINZ, B., STRACHAN, I. (2016). A fundamental relationship between intraventricular conduction and heart rate". Journal of Electrocardiology ,Vol. 49, No. 3.

[MD SAUDE, 2019] - MD. Saúde - Website disponível em "https://www.mdsaude.com/cardiologia/prolapso-valvula-mitral/ ", acessado em 24 de novembro de 2019.

[METZ, 1978] - METZ, C. E. (1978). "Basic Principles of ROC Analysis". Seminars in Nuclear Medicine, Vol. 8, No. 4.

[MIT-BIH, 2000a] - (2000). "The MIT-BIH Arrhythmia Database". Website disponível em "http://www.physionet.org/physiobank/database/mitdb/", Acessado em 18 de maio de 2019.

[MIT-BIH, 2000b] - (2000). "The MIT-BIH Normal Sinus Rhythm Database". Website disponível em "http://www.physionet.zorg/physiobank/database/nsrdb/", Acessado em 18 de maio de 2019.

[MOHSIN, 2017] - MOHSIN, J. (2017). "Two Factor Vs Multi-Factor, an Authentication Battle in Mobile Cloud Computing Environments". International Conference on Future Networks and Distributed Systems Proceedings, No. 39.

[PAWLE, 2013] - PAWLE, A. A., PAWAR V. P. (2013) "Face Recognition System (FRS) on Cloud Computing for User Authentication". IJSCE - International Journal of Soft Computing and Engineering, Vol. 3, No. 4.

[POURSHAHID, 2007] - POURSHAHID, A., TRAN, T. (2007). "Modeling Trust in E-Commerce: An Approach Based on User Requirements". ICEC 2007 - IEEE International Conference on Electronic E-commerce.

[PUSARA, 2004] - PUSARA, M., BRODLEY, C. E. (2004). "User ReAuthentication via Mouse Movements". 2004 ACM workshop on Visualization and data mining for computer security.

[QARDIO CORE, 2019] - "Qardicore". Website disponível em "https://www.getqardio.com/qardiocore-wearable-ecg-ekg-monitor-iphone/", Acessado em 11 de julho de 2019.

[RAHAL, 2007] - RAHAL, S. M., ABOALSAMAH, H. A., MUTEB, K. N. (2007). "Multimodal Biometric Authentication System – MBAS". ICTTA '06 - 2nd Information and Communication Technologies.

[RAJA, 2013] - RAJA, A. Y., PERUMAL S. A. (2013) "Effective Method of Web Site Authentication Using Finger Print Verification". International Journal of Computer and Electrical Engineering, Vol. 5, No. 6.

[REZGUI, 2016] REZGUI, D., LACHIRI, Z. (2016), "ECG Biometric Recognition Using SVM-Based Approach". Institute of Electrical Engineers of Japan - Transactions on Electrical and Electronic Engineering, Vol. 11, No. S1.

[RIERA, 2009] - RIERA, A., SORIAFRISCH, A., CAPARRINI, M., CESTER, I., RUFFINI G. (2009) "Multimodal Physiological Biometrics Authentication". Capítulo do Livro "Biometrics: Theory, Methods, and Applications", John Wiley & Sons.

[RYAN, 2013] - RYAN, M. D. (2013). "Cloud Computing security: The scientific challenge, and a survey of solutions". Journal of Systems and Software, Vol. 86, No. 9.

[SADDIK, 2017] - EL SADDIK, A., FALCONI, J. S. A., AL OSMAN, H. (2017). "Elecctrocardiogram (ECG) Biometric Authentication". US PATENT.

[SATYANARAYANAN, 2009] - SATYANARAYANAN, M., BAHL, P., CACERES, R., DAVIES, N. (2009). "The case for vm-based cloudlets in mobile computing", IEEE Pervasive Computing, Vol. 8, No. 4.

[SAUVER, 2009] - SAUVER, J. S. (2009). "Passwords". NWACC - Northwest Academic Computing Consortium Security Meeting.

[SCHNEIER, 1996] - SCHNEIER, B. (1996). "Applied Cryptography: Protocols, Algorithms, and Source Code in C". Livro, Segunda edição, John Wiley & Sons.

[SHIRAZ, 2013] - SHIRAZ, M., GANI, A., KHOKHAR, R. H., BUYYA, R. (2013). "A Review on Distributed Application Processing Frameworks in Smart Mobile Devices for Mobile Cloud Computing", IEEE Communications Surveys & Tutorials, Vol. 15, No. 3.

[SIMEONE, 2017] - SIMEONE. O. (2017). "A Brief Introduction to Machine Learning for Engineers". Computer Science, Mathematics, Foundations and Trends in Signal Processing 2017.

[SINGH, 2012] - SINGH, Y. N., SINGH, S. K. (2012). "Evaluation of Electrocardiogram for Biometric Authentication.(Report)." Journal of Information Security, Vol. 3, No. 1.

[SOUZA, 2019] - SOUZA, P. H. B. (2019). "Método para estimação da frequência cardíaca e variabilidade cardíaca com base em fotopletismografia por vídeo". Dissertação de Mestrado em Engenharia Biomédica, Programa de Pós-Graduação em Engenharia Biomédica, Faculdade Gama, Universidade de Brasília, Brasília, DF, Publicação 109A/2019.

[STOJMENOVIC, 2012] - STOJMENOVIC, M. (2012) "Mobile Cloud Computing for Biometric Applications". Proceedings of 15th International Conference on NetworkBased Information Systems.

[SUO, 2013] - SUO, H., LIU, Z., WAN, J., ZHOU, K., (2013). "Security and privacy in mobile Cloud Computing,", IWCMC - 9th International Wireless Communications and Mobile Computing Conference.

[SYTA, 2010] - SYTA, E., KURKOVSKY, S., CASANO B. (2010). "RFID-Based Authentication Middleware for Mobile Devices". HICSS - 43rd Hawaii International Conference on System Sciences.

[TADDEI, 1992] - TADDEI, A. (1992). “The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography”. European Heart Journal, Vol. 13, No. 9.

[THOMPSON, 2002] - THOMPSON, M. R. , OLSON, D., COWLES, R., MULLEN, S., HELM, M. (2002). "CA-based Trust Model for Grid Authentication and Identity Delegation". Website disponivel em "http://www.gridcp.es.net/Documents/GGF6/TrustModel-final.pdf", Acessado em 13 de setembro de 2019.

[TORTORA, 2014] - TORTORA, G. J., DERRICKSON, B. (2014). "Principles of anatomy and physiology". Livro, décima quarta edição. John Wiley & Sons.

[TSAI, 2015] - TSAI, J., LO, N. (2015). "A Privacy-Aware Authentication Scheme for Distributed Mobile Cloud Computing Services". IEEE Systems Journal, Vol. 9, No. 3.

[VAPNIK, 2004] - VAPNIK, V., CHERVONENKIS, A. (2004). "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities". Capítulo do Livro "Measures of Complexity", Springer.

[WADHWA, 2014] - WADHWA, A., GUPTA V. K. (2014) "Framework for User Authenticity and Access Control Security over a Cloud". IJCSE - International Journal on Computer Science and Engineering, Vol. 6, No.4.

[WAN, 2013] WAN, J., LIU, Z., ZHOU, K., LU, R. (2013). "Mobile Cloud Computing: Application Scenarios and Service Models", IWCMC - 9th International Wireless Communications and Mobile Computing Conference.

[WANG, 2007] - WANG, Y. , VASSILEVA, J. (2007). "A Review on Trust and Reputation for Web Service Selection". ICDCSW'07 - 27th International Conference on Distributed Computing Systems Workshops .

[WANG, 2011] - WANG, P., KU, C. C., WANG T. C. (2011) "A New Fingerprint Authentication Scheme based on Secret-splitting for Cloud Computing Security". Recent Application in Biometrics, Disponível em "http://www.intechopen.com/books/recent-application-in-biometrics/a-new-fingerprint-authentication-scheme-based-on-secret-splitting-for-cloud-computing-security", Acessado em 21 de agosto de 2019.

[XIONG, 2017] - XIONG, L. (2017). "An Enhanced Privacy-Aware Authentication Scheme for Distributed Mobile Cloud Computing Services". KSII Transactions on Internet and Information Systems, Vol. 11, No. 12.

[YE, 2015] - YE, P., YU, M., WU M. (2015) "Implementation: Mobile Face Identity Authentication System on Android Platforms". International Journal of Security and Its Applications, Vol.9, No.1.

[ZHANG, 2010] - ZHANG, X., JEONG, S., KUNJITHAPATHAM, A., GIBBS, S. (2010). "Towards an elastic application model for augmenting computing capabilities of mobile platforms". Capítulo do Livro "Mobile wireless middleware, operating systems, and applications". Primeira Edição. Springer.

[ZHOU, 2012a] - ZHOU, Z., HUANG, D. (2012). "Efficient and Secure Data Storage Operations for Mobile Cloud Computing". 8th International Conference and Wworkshop on Systems Virtualiztion Management.

[ZHOU, 2012b] - ZHOU, Z. (2012)."Ensemble Methods - Foundations and Algorithms". Livro, Primeira Edição, Chapman and Hall - CRC.

[ZISSIS, 2012] - ZISSIS, D., LEKKAS, D. (2012). "Addressing Cloud Computing security issues." Future Generation Computer Systems, Vol. 28.

[ZKIK, 2016] - ZKIK, K. (2016). "Secure Scheme on Mobile Multi Cloud Computing Based on Homomorphic Encryption". ICEMIS - International Conference On Engineering &Amp.

APÊNDICES

APÊNDICE A - TABELAS COMPLETAS DOS VALORES DAS MÉTRICAS DE ANÁLISE REFERENTES AOS CLASSIFICADORES

Tabela A.1 - SVM - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando todos os usuários (ordenado com base em NC_u).

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,820 ± 0,241	0,640 ± 0,482	0,640 ± 0,482	0,640 ± 0,482
4	69	0,495 ± 0,050	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	12	0,805 ± 0,245	0,610 ± 0,490	0,610 ± 0,490	0,610 ± 0,490
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
6	53	0,660 ± 0,149	0,600 ± 0,492	0,325 ± 0,288	0,417 ± 0,349
7	81	0,753 ± 0,176	0,775 ± 0,411	0,520 ± 0,340	0,605 ± 0,345
11	52	0,632 ± 0,137	0,603 ± 0,469	0,303 ± 0,264	0,390 ± 0,317
11	66	0,768 ± 0,138	0,911 ± 0,266	0,553 ± 0,269	0,664 ± 0,254
12	48	0,670 ± 0,121	0,818 ± 0,348	0,380 ± 0,232	0,496 ± 0,254
12	78	0,809 ± 0,109	0,990 ± 0,100	0,618 ± 0,217	0,739 ± 0,188
12	91	0,728 ± 0,110	0,940 ± 0,228	0,460 ± 0,212	0,598 ± 0,219
13	76	0,499 ± 0,013	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
16	64	0,922 ± 0,101	0,971 ± 0,075	0,874 ± 0,183	0,908 ± 0,131
17	43	0,899 ± 0,088	0,989 ± 0,045	0,810 ± 0,174	0,878 ± 0,119
18	45	0,999 ± 0,010	1,000 ± 0,000	0,998 ± 0,020	0,999 ± 0,011
25	107	0,841 ± 0,087	0,917 ± 0,106	0,766 ± 0,164	0,820 ± 0,109
33	44	0,957 ± 0,057	0,965 ± 0,060	0,953 ± 0,093	0,955 ± 0,064
35	10	0,791 ± 0,076	0,917 ± 0,094	0,647 ± 0,139	0,749 ± 0,107
36	35	0,807 ± 0,064	0,921 ± 0,085	0,680 ± 0,114	0,774 ± 0,085
37	8	0,973 ± 0,039	0,983 ± 0,037	0,963 ± 0,063	0,972 ± 0,042
38	3	0,897 ± 0,058	0,940 ± 0,067	0,854 ± 0,102	0,890 ± 0,068
39	47	0,928 ± 0,053	0,962 ± 0,056	0,896 ± 0,094	0,924 ± 0,059
42	98	0,723 ± 0,076	0,863 ± 0,118	0,534 ± 0,130	0,651 ± 0,116
43	28	0,937 ± 0,046	0,975 ± 0,041	0,899 ± 0,085	0,933 ± 0,051
43	61	0,910 ± 0,060	0,910 ± 0,079	0,919 ± 0,084	0,911 ± 0,061
46	94	0,877 ± 0,050	0,966 ± 0,048	0,784 ± 0,092	0,862 ± 0,061
49	99	0,950 ± 0,041	0,986 ± 0,033	0,914 ± 0,074	0,947 ± 0,045
53	59	0,903 ± 0,056	0,984 ± 0,033	0,821 ± 0,107	0,891 ± 0,069
56	105	0,887 ± 0,051	0,948 ± 0,056	0,822 ± 0,089	0,878 ± 0,060
65	16	0,934 ± 0,036	0,962 ± 0,040	0,904 ± 0,057	0,931 ± 0,038
82	36	0,925 ± 0,036	0,944 ± 0,047	0,907 ± 0,051	0,924 ± 0,037
82	79	0,946 ± 0,026	0,981 ± 0,025	0,911 ± 0,048	0,944 ± 0,029
87	2	0,830 ± 0,055	0,835 ± 0,064	0,828 ± 0,084	0,829 ± 0,057
89	37	0,835 ± 0,048	0,937 ± 0,051	0,720 ± 0,090	0,811 ± 0,063
97	39	0,829 ± 0,048	0,859 ± 0,056	0,790 ± 0,082	0,820 ± 0,055
102	83	0,859 ± 0,037	0,889 ± 0,051	0,824 ± 0,059	0,854 ± 0,039

103	27	0,952	±	0,022	0,964	±	0,033	0,941	±	0,035	0,951	± 0,022
104	77	0,882	±	0,039	0,889	±	0,053	0,876	±	0,056	0,881	± 0,039
109	62	0,888	±	0,037	0,891	±	0,051	0,887	±	0,050	0,888	± 0,037
117	87	0,820	±	0,042	0,845	±	0,047	0,787	±	0,076	0,812	± 0,049
120	24	0,908	±	0,036	0,893	±	0,048	0,931	±	0,042	0,911	± 0,035
121	54	0,945	±	0,026	0,917	±	0,040	0,981	±	0,026	0,947	± 0,024
123	86	0,874	±	0,036	0,871	±	0,052	0,883	±	0,053	0,876	± 0,036
127	4	0,811	±	0,045	0,815	±	0,053	0,809	±	0,064	0,810	± 0,046
128	31	0,921	±	0,032	0,907	±	0,040	0,941	±	0,039	0,923	± 0,031
136	6	0,816	±	0,036	0,789	±	0,045	0,868	±	0,048	0,825	± 0,033
143	25	0,927	±	0,027	0,959	±	0,033	0,893	±	0,049	0,924	± 0,030
145	33	0,955	±	0,024	0,925	±	0,038	0,991	±	0,017	0,956	± 0,022
147	85	0,916	±	0,034	0,880	±	0,042	0,966	±	0,034	0,920	± 0,032
148	68	0,929	±	0,023	0,924	±	0,033	0,937	±	0,034	0,930	± 0,023
162	97	0,936	±	0,024	0,900	±	0,034	0,982	±	0,020	0,939	± 0,022
170	19	0,963	±	0,017	0,950	±	0,028	0,979	±	0,021	0,964	± 0,017
173	102	0,895	±	0,030	0,906	±	0,043	0,885	±	0,043	0,894	± 0,031
175	21	0,935	±	0,024	0,940	±	0,031	0,931	±	0,036	0,935	± 0,024
179	15	0,916	±	0,024	0,900	±	0,037	0,938	±	0,032	0,918	± 0,023
185	5	0,916	±	0,023	0,902	±	0,029	0,935	±	0,033	0,918	± 0,023
186	80	0,954	±	0,019	0,952	±	0,026	0,957	±	0,030	0,954	± 0,019
196	70	0,921	±	0,023	0,929	±	0,035	0,913	±	0,033	0,920	± 0,023
200	82	0,897	±	0,031	0,901	±	0,037	0,893	±	0,042	0,896	± 0,032
207	65	0,912	±	0,024	0,886	±	0,032	0,946	±	0,029	0,915	± 0,023
212	90	0,850	±	0,026	0,801	±	0,033	0,933	±	0,030	0,861	± 0,023
214	11	0,935	±	0,022	0,922	±	0,031	0,952	±	0,025	0,936	± 0,021
215	60	0,912	±	0,023	0,886	±	0,034	0,947	±	0,025	0,915	± 0,022
248	84	0,950	±	0,016	0,955	±	0,023	0,947	±	0,026	0,950	± 0,016
266	49	0,951	±	0,017	0,922	±	0,026	0,985	±	0,021	0,952	± 0,016
270	9	0,951	±	0,016	0,942	±	0,022	0,962	±	0,022	0,952	± 0,016
280	46	0,981	±	0,012	0,979	±	0,014	0,983	±	0,018	0,981	± 0,012
281	20	0,987	±	0,009	0,987	±	0,013	0,987	±	0,012	0,987	± 0,009
290	34	0,927	±	0,021	0,892	±	0,031	0,972	±	0,019	0,930	± 0,019
292	92	0,963	±	0,014	0,943	±	0,023	0,987	±	0,014	0,964	± 0,013
293	17	0,971	±	0,013	0,972	±	0,016	0,970	±	0,021	0,971	± 0,013
300	1	0,937	±	0,018	0,911	±	0,028	0,969	±	0,018	0,939	± 0,016
300	7	0,961	±	0,015	0,939	±	0,025	0,987	±	0,014	0,962	± 0,015
300	13	0,908	±	0,021	0,877	±	0,029	0,950	±	0,023	0,912	± 0,019
300	14	0,895	±	0,018	0,855	±	0,027	0,952	±	0,025	0,901	± 0,016
300	18	0,952	±	0,017	0,928	±	0,026	0,980	±	0,015	0,953	± 0,016
300	22	0,956	±	0,015	0,946	±	0,021	0,969	±	0,019	0,957	± 0,015
300	23	0,938	±	0,017	0,917	±	0,027	0,964	±	0,021	0,940	± 0,016
300	26	0,905	±	0,020	0,859	±	0,029	0,970	±	0,016	0,911	± 0,018
300	29	0,843	±	0,027	0,809	±	0,031	0,901	±	0,031	0,852	± 0,024
300	30	0,854	±	0,026	0,800	±	0,032	0,947	±	0,027	0,867	± 0,022

300	32	0,937	±	0,018	0,918	±	0,025	0,960	±	0,019	0,938	±	0,017
300	38	0,927	±	0,017	0,889	±	0,027	0,978	±	0,019	0,931	±	0,015
300	40	0,890	±	0,023	0,857	±	0,032	0,937	±	0,024	0,895	±	0,021
300	41	0,936	±	0,017	0,907	±	0,027	0,972	±	0,019	0,938	±	0,016
300	42	0,972	±	0,013	0,962	±	0,020	0,983	±	0,013	0,972	±	0,013
300	50	0,929	±	0,020	0,911	±	0,028	0,952	±	0,023	0,931	±	0,019
300	51	0,959	±	0,014	0,940	±	0,024	0,983	±	0,014	0,960	±	0,013
300	55	0,967	±	0,011	0,948	±	0,019	0,990	±	0,012	0,968	±	0,011
300	56	0,949	±	0,016	0,934	±	0,024	0,968	±	0,017	0,950	±	0,015
300	57	0,991	±	0,007	0,983	±	0,013	1,000	±	0,000	0,992	±	0,007
300	58	0,967	±	0,015	0,959	±	0,022	0,976	±	0,017	0,968	±	0,014
300	63	0,959	±	0,014	0,947	±	0,022	0,973	±	0,016	0,960	±	0,014
300	71	0,962	±	0,013	0,940	±	0,021	0,988	±	0,010	0,963	±	0,012
300	72	0,959	±	0,013	0,955	±	0,021	0,964	±	0,018	0,959	±	0,013
300	73	0,993	±	0,007	0,997	±	0,005	0,989	±	0,012	0,993	±	0,007
300	74	0,978	±	0,012	0,965	±	0,020	0,992	±	0,009	0,978	±	0,012
300	75	0,940	±	0,017	0,920	±	0,026	0,964	±	0,017	0,942	±	0,016
300	88	0,957	±	0,015	0,931	±	0,026	0,989	±	0,012	0,959	±	0,014
300	89	0,953	±	0,014	0,950	±	0,021	0,957	±	0,019	0,953	±	0,014
300	93	0,972	±	0,012	0,952	±	0,021	0,994	±	0,009	0,972	±	0,012
300	95	0,960	±	0,013	0,958	±	0,018	0,962	±	0,019	0,960	±	0,013
300	96	0,991	±	0,008	0,991	±	0,010	0,991	±	0,013	0,991	±	0,008
300	100	0,995	±	0,006	0,995	±	0,008	0,996	±	0,009	0,995	±	0,006
300	101	0,949	±	0,016	0,960	±	0,021	0,938	±	0,025	0,949	±	0,016
300	103	0,914	±	0,020	0,880	±	0,028	0,959	±	0,021	0,918	±	0,019
300	106	0,984	±	0,008	0,980	±	0,014	0,988	±	0,010	0,984	±	0,008
300	108	0,991	±	0,007	0,983	±	0,014	1,000	±	0,000	0,992	±	0,007

Tabela A.2 - *AdaBoost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando todos os usuários (ordenado com base em NC_u).

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
4	69	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	12	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
6	53	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
7	81	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	66	0,608 ± 0,143	0,420 ± 0,496	0,217 ± 0,286	0,277 ± 0,344
12	48	0,505 ± 0,025	0,040 ± 0,197	0,010 ± 0,049	0,016 ± 0,079
12	78	0,501 ± 0,013	0,010 ± 0,100	0,003 ± 0,025	0,004 ± 0,040
12	91	0,510 ± 0,034	0,080 ± 0,273	0,020 ± 0,068	0,032 ± 0,109
13	76	0,519 ± 0,051	0,150 ± 0,359	0,040 ± 0,099	0,063 ± 0,152
16	64	0,608 ± 0,132	0,490 ± 0,502	0,216 ± 0,263	0,287 ± 0,323
17	43	0,549 ± 0,116	0,190 ± 0,394	0,098 ± 0,232	0,122 ± 0,272
18	45	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
25	107	0,729 ± 0,095	1,000 ± 0,000	0,459 ± 0,191	0,605 ± 0,183
33	44	0,706 ± 0,083	0,990 ± 0,100	0,411 ± 0,167	0,562 ± 0,180
35	10	0,660 ± 0,076	1,000 ± 0,000	0,320 ± 0,151	0,466 ± 0,170
36	35	0,690 ± 0,062	0,990 ± 0,100	0,379 ± 0,123	0,538 ± 0,138
37	8	0,733 ± 0,192	0,630 ± 0,485	0,466 ± 0,383	0,529 ± 0,418
38	3	0,790 ± 0,080	1,000 ± 0,000	0,580 ± 0,159	0,721 ± 0,137
39	47	0,816 ± 0,083	1,000 ± 0,000	0,632 ± 0,167	0,761 ± 0,132
42	98	0,504 ± 0,013	0,110 ± 0,314	0,008 ± 0,024	0,016 ± 0,045
43	28	0,889 ± 0,073	0,989 ± 0,100	0,779 ± 0,146	0,866 ± 0,120
43	61	0,678 ± 0,067	1,000 ± 0,000	0,357 ± 0,134	0,512 ± 0,149
46	94	0,706 ± 0,058	1,000 ± 0,000	0,413 ± 0,117	0,575 ± 0,118
49	99	0,904 ± 0,101	0,960 ± 0,197	0,808 ± 0,202	0,873 ± 0,194
53	59	0,833 ± 0,060	1,000 ± 0,000	0,666 ± 0,119	0,793 ± 0,091
56	105	0,732 ± 0,060	1,000 ± 0,000	0,464 ± 0,121	0,624 ± 0,115
65	16	0,806 ± 0,059	1,000 ± 0,000	0,613 ± 0,118	0,753 ± 0,096
82	36	0,895 ± 0,041	1,000 ± 0,000	0,790 ± 0,082	0,880 ± 0,053
82	79	0,934 ± 0,031	1,000 ± 0,000	0,868 ± 0,061	0,928 ± 0,035
87	2	0,782 ± 0,044	1,000 ± 0,000	0,564 ± 0,088	0,717 ± 0,074
89	37	0,811 ± 0,043	1,000 ± 0,000	0,622 ± 0,085	0,764 ± 0,065
97	39	0,539 ± 0,024	0,885 ± 0,317	0,079 ± 0,048	0,142 ± 0,083
102	83	0,663 ± 0,034	0,997 ± 0,016	0,327 ± 0,067	0,488 ± 0,076
103	27	0,937 ± 0,056	0,999 ± 0,007	0,874 ± 0,111	0,928 ± 0,069
104	77	0,758 ± 0,044	0,999 ± 0,009	0,516 ± 0,087	0,676 ± 0,077
109	62	0,744 ± 0,044	0,998 ± 0,011	0,490 ± 0,088	0,652 ± 0,081
117	87	0,566 ± 0,027	0,988 ± 0,103	0,133 ± 0,053	0,230 ± 0,082

120	24	0,693	±	0,035	0,996	±	0,015	0,387	±	0,071	0,553	± 0,076
121	54	0,836	±	0,037	0,998	±	0,008	0,673	±	0,074	0,801	± 0,053
123	86	0,643	±	0,033	0,998	±	0,013	0,286	±	0,067	0,441	± 0,081
127	4	0,635	±	0,036	1,000	±	0,000	0,271	±	0,072	0,421	± 0,091
128	31	0,920	±	0,027	1,000	±	0,004	0,841	±	0,053	0,912	± 0,032
136	6	0,761	±	0,031	1,000	±	0,005	0,523	±	0,062	0,685	± 0,054
143	25	0,907	±	0,025	1,000	±	0,003	0,813	±	0,049	0,896	± 0,030
145	33	0,810	±	0,037	0,999	±	0,006	0,621	±	0,075	0,763	± 0,059
147	85	0,819	±	0,039	1,000	±	0,000	0,638	±	0,078	0,776	± 0,059
148	68	0,816	±	0,038	0,998	±	0,008	0,632	±	0,076	0,771	± 0,058
162	97	0,920	±	0,025	1,000	±	0,003	0,840	±	0,049	0,912	± 0,030
170	19	0,929	±	0,025	0,999	±	0,005	0,859	±	0,050	0,923	± 0,029
173	102	0,866	±	0,030	0,999	±	0,004	0,732	±	0,060	0,844	± 0,040
175	21	0,917	±	0,021	0,999	±	0,006	0,834	±	0,042	0,909	± 0,025
179	15	0,800	±	0,030	0,997	±	0,008	0,602	±	0,061	0,749	± 0,047
185	5	0,835	±	0,035	0,999	±	0,006	0,670	±	0,070	0,800	± 0,051
186	80	0,836	±	0,034	0,997	±	0,008	0,674	±	0,068	0,803	± 0,049
196	70	0,872	±	0,024	0,999	±	0,004	0,744	±	0,047	0,852	± 0,031
200	82	0,877	±	0,030	0,999	±	0,006	0,755	±	0,059	0,858	± 0,039
207	65	0,808	±	0,033	0,998	±	0,006	0,617	±	0,065	0,761	± 0,050
212	90	0,744	±	0,031	0,997	±	0,009	0,489	±	0,061	0,654	± 0,055
214	11	0,898	±	0,025	0,999	±	0,004	0,797	±	0,051	0,886	± 0,032
215	60	0,893	±	0,026	1,000	±	0,003	0,785	±	0,052	0,879	± 0,033
248	84	0,953	±	0,016	1,000	±	0,001	0,907	±	0,032	0,951	± 0,018
266	49	0,966	±	0,015	0,999	±	0,003	0,932	±	0,031	0,964	± 0,017
270	9	0,975	±	0,012	1,000	±	0,001	0,950	±	0,025	0,974	± 0,013
280	46	0,746	±	0,247	0,500	±	0,502	0,492	±	0,495	0,496	± 0,499
281	20	0,985	±	0,009	1,000	±	0,001	0,970	±	0,018	0,985	± 0,010
290	34	0,969	±	0,012	1,000	±	0,000	0,938	±	0,025	0,968	± 0,013
292	92	0,978	±	0,014	1,000	±	0,002	0,955	±	0,027	0,977	± 0,014
293	17	0,966	±	0,013	1,000	±	0,002	0,932	±	0,026	0,964	± 0,014
300	1	0,924	±	0,017	0,999	±	0,003	0,849	±	0,034	0,917	± 0,020
300	7	0,988	±	0,010	1,000	±	0,000	0,977	±	0,020	0,988	± 0,010
300	13	0,878	±	0,021	0,999	±	0,004	0,757	±	0,043	0,861	± 0,028
300	14	0,850	±	0,023	0,998	±	0,005	0,701	±	0,046	0,823	± 0,032
300	18	0,894	±	0,021	0,998	±	0,006	0,789	±	0,042	0,881	± 0,027
300	22	0,915	±	0,021	0,998	±	0,006	0,832	±	0,041	0,907	± 0,025
300	23	0,946	±	0,015	1,000	±	0,002	0,892	±	0,030	0,943	± 0,017
300	26	0,931	±	0,021	0,999	±	0,004	0,863	±	0,041	0,926	± 0,024
300	29	0,816	±	0,023	0,998	±	0,005	0,633	±	0,046	0,774	± 0,035
300	30	0,874	±	0,022	0,998	±	0,005	0,750	±	0,043	0,855	± 0,028
300	32	0,881	±	0,021	0,999	±	0,004	0,762	±	0,043	0,864	± 0,028
300	38	0,944	±	0,017	0,998	±	0,005	0,889	±	0,034	0,940	± 0,019
300	40	0,765	±	0,027	0,996	±	0,009	0,533	±	0,053	0,693	± 0,045
300	41	0,987	±	0,010	1,000	±	0,001	0,975	±	0,020	0,987	± 0,011

300	42	0,967	±	0,014	1,000	±	0,002	0,934	±	0,028	0,965	± 0,015
300	50	0,957	±	0,015	1,000	±	0,002	0,914	±	0,029	0,955	± 0,016
300	51	0,974	±	0,011	1,000	±	0,002	0,948	±	0,023	0,973	± 0,012
300	55	0,981	±	0,011	1,000	±	0,003	0,962	±	0,022	0,980	± 0,012
300	56	0,938	±	0,018	0,999	±	0,004	0,878	±	0,035	0,934	± 0,020
300	57	0,988	±	0,009	1,000	±	0,000	0,975	±	0,018	0,987	± 0,009
300	58	0,937	±	0,016	0,999	±	0,004	0,875	±	0,033	0,933	± 0,019
300	63	0,959	±	0,013	1,000	±	0,002	0,918	±	0,026	0,957	± 0,014
300	71	0,944	±	0,017	0,999	±	0,004	0,889	±	0,033	0,940	± 0,019
300	72	0,975	±	0,013	1,000	±	0,000	0,950	±	0,025	0,974	± 0,013
300	73	0,590	±	0,186	0,190	±	0,394	0,179	±	0,372	0,184	± 0,383
300	74	0,985	±	0,015	1,000	±	0,001	0,969	±	0,030	0,984	± 0,017
300	75	0,910	±	0,022	0,998	±	0,005	0,821	±	0,045	0,900	± 0,027
300	88	0,972	±	0,011	1,000	±	0,002	0,944	±	0,022	0,971	± 0,012
300	89	0,957	±	0,015	1,000	±	0,001	0,914	±	0,030	0,955	± 0,016
300	93	0,983	±	0,010	1,000	±	0,002	0,965	±	0,021	0,982	± 0,011
300	95	0,934	±	0,017	0,999	±	0,002	0,868	±	0,033	0,929	± 0,019
300	96	0,992	±	0,007	1,000	±	0,001	0,984	±	0,014	0,992	± 0,007
300	100	0,873	±	0,211	0,760	±	0,429	0,746	±	0,421	0,753	± 0,425
300	101	0,951	±	0,015	1,000	±	0,002	0,902	±	0,029	0,948	± 0,016
300	103	0,848	±	0,025	0,997	±	0,007	0,699	±	0,050	0,821	± 0,035
300	106	0,829	±	0,232	0,670	±	0,472	0,658	±	0,464	0,663	± 0,468
300	108	0,992	±	0,006	0,999	±	0,002	0,985	±	0,013	0,992	± 0,007

Tabela A.3 - *Robust Boost* - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando todos os usuários (ordenado com base em NC_u).

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,560 ± 0,163	0,120 ± 0,327	0,120 ± 0,327	0,120 ± 0,327
4	69	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	12	0,610 ± 0,208	0,220 ± 0,416	0,220 ± 0,416	0,220 ± 0,416
5	67	0,555 ± 0,157	0,110 ± 0,314	0,110 ± 0,314	0,110 ± 0,314
6	53	0,505 ± 0,035	0,020 ± 0,141	0,010 ± 0,070	0,013 ± 0,094
7	81	0,633 ± 0,148	0,480 ± 0,502	0,265 ± 0,297	0,337 ± 0,359
11	52	0,503 ± 0,023	0,020 ± 0,141	0,007 ± 0,047	0,010 ± 0,070
11	66	0,665 ± 0,145	0,650 ± 0,479	0,330 ± 0,290	0,424 ± 0,340
12	48	0,509 ± 0,032	0,070 ± 0,256	0,018 ± 0,064	0,028 ± 0,103
12	78	0,763 ± 0,122	0,940 ± 0,239	0,525 ± 0,245	0,651 ± 0,240
12	91	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
13	76	0,541 ± 0,071	0,290 ± 0,456	0,083 ± 0,142	0,126 ± 0,206
16	64	0,753 ± 0,131	0,970 ± 0,171	0,506 ± 0,261	0,631 ± 0,241
17	43	0,786 ± 0,134	1,000 ± 0,000	0,572 ± 0,268	0,690 ± 0,226
18	45	1,000 ± 0,000	1,000 ± 0,000	1,000 ± 0,000	1,000 ± 0,000
25	107	0,694 ± 0,102	0,980 ± 0,141	0,387 ± 0,203	0,528 ± 0,210
33	44	0,694 ± 0,076	1,000 ± 0,000	0,387 ± 0,152	0,541 ± 0,163
35	10	0,636 ± 0,070	0,970 ± 0,171	0,272 ± 0,140	0,409 ± 0,173
36	35	0,670 ± 0,066	0,980 ± 0,141	0,340 ± 0,133	0,492 ± 0,154
37	8	0,886 ± 0,075	1,000 ± 0,000	0,773 ± 0,149	0,863 ± 0,103
38	3	0,781 ± 0,081	1,000 ± 0,000	0,563 ± 0,162	0,706 ± 0,142
39	47	0,792 ± 0,090	1,000 ± 0,000	0,584 ± 0,180	0,721 ± 0,148
42	98	0,505 ± 0,015	0,140 ± 0,349	0,012 ± 0,030	0,021 ± 0,054
43	28	0,893 ± 0,054	0,999 ± 0,009	0,788 ± 0,107	0,877 ± 0,072
43	61	0,668 ± 0,064	1,000 ± 0,000	0,335 ± 0,128	0,488 ± 0,150
46	94	0,696 ± 0,057	1,000 ± 0,000	0,393 ± 0,115	0,554 ± 0,120
49	99	0,912 ± 0,050	0,999 ± 0,007	0,825 ± 0,100	0,900 ± 0,062
53	59	0,811 ± 0,062	1,000 ± 0,000	0,623 ± 0,124	0,760 ± 0,097
56	105	0,718 ± 0,065	1,000 ± 0,000	0,436 ± 0,131	0,596 ± 0,127
65	16	0,788 ± 0,057	1,000 ± 0,000	0,577 ± 0,114	0,725 ± 0,096
82	36	0,876 ± 0,039	1,000 ± 0,000	0,752 ± 0,079	0,856 ± 0,052
82	79	0,934 ± 0,029	1,000 ± 0,000	0,869 ± 0,058	0,929 ± 0,033
87	2	0,757 ± 0,057	0,999 ± 0,010	0,515 ± 0,114	0,671 ± 0,105
89	37	0,781 ± 0,056	0,999 ± 0,006	0,563 ± 0,111	0,713 ± 0,100
97	39	0,539 ± 0,029	0,859 ± 0,338	0,081 ± 0,057	0,144 ± 0,095
102	83	0,628 ± 0,063	0,909 ± 0,287	0,257 ± 0,127	0,392 ± 0,176
103	27	0,955 ± 0,028	1,000 ± 0,003	0,911 ± 0,055	0,953 ± 0,031
104	77	0,739 ± 0,046	0,999 ± 0,009	0,478 ± 0,093	0,641 ± 0,084
109	62	0,723 ± 0,054	0,988 ± 0,100	0,447 ± 0,109	0,609 ± 0,113
117	87	0,549 ± 0,028	0,928 ± 0,257	0,097 ± 0,055	0,173 ± 0,091

120	24	0,684	±	0,046	0,996	±	0,017	0,369	±	0,091	0,532	± 0,098
121	54	0,806	±	0,051	0,998	±	0,011	0,613	±	0,103	0,754	± 0,086
123	86	0,626	±	0,055	0,975	±	0,142	0,254	±	0,110	0,391	± 0,147
127	4	0,621	±	0,043	0,987	±	0,101	0,244	±	0,088	0,383	± 0,120
128	31	0,909	±	0,030	0,999	±	0,004	0,819	±	0,061	0,899	± 0,037
136	6	0,735	±	0,049	0,998	±	0,020	0,471	±	0,096	0,634	± 0,097
143	25	0,900	±	0,028	1,000	±	0,000	0,801	±	0,056	0,888	± 0,035
145	33	0,797	±	0,042	0,999	±	0,007	0,595	±	0,085	0,742	± 0,068
147	85	0,794	±	0,051	0,999	±	0,004	0,588	±	0,103	0,735	± 0,084
148	68	0,764	±	0,055	0,998	±	0,010	0,528	±	0,110	0,684	± 0,099
162	97	0,900	±	0,031	0,999	±	0,004	0,801	±	0,062	0,888	± 0,039
170	19	0,917	±	0,029	1,000	±	0,003	0,834	±	0,057	0,908	± 0,034
173	102	0,854	±	0,032	0,999	±	0,005	0,708	±	0,065	0,827	± 0,045
175	21	0,906	±	0,023	0,999	±	0,004	0,812	±	0,045	0,895	± 0,028
179	15	0,785	±	0,033	0,995	±	0,013	0,573	±	0,066	0,725	± 0,053
185	5	0,794	±	0,053	0,997	±	0,012	0,589	±	0,104	0,735	± 0,091
186	80	0,807	±	0,042	0,996	±	0,012	0,617	±	0,082	0,759	± 0,065
196	70	0,852	±	0,032	0,998	±	0,007	0,705	±	0,063	0,825	± 0,044
200	82	0,846	±	0,044	0,999	±	0,005	0,692	±	0,088	0,814	± 0,068
207	65	0,742	±	0,049	0,996	±	0,012	0,485	±	0,098	0,647	± 0,090
212	90	0,708	±	0,054	0,985	±	0,100	0,418	±	0,108	0,579	± 0,119
214	11	0,868	±	0,043	0,999	±	0,005	0,737	±	0,086	0,845	± 0,063
215	60	0,877	±	0,028	0,999	±	0,004	0,754	±	0,055	0,858	± 0,037
248	84	0,950	±	0,017	1,000	±	0,002	0,901	±	0,035	0,947	± 0,019
266	49	0,954	±	0,019	0,999	±	0,004	0,909	±	0,037	0,951	± 0,021
270	9	0,962	±	0,017	1,000	±	0,002	0,925	±	0,035	0,960	± 0,019
280	46	0,986	±	0,011	1,000	±	0,002	0,973	±	0,022	0,986	± 0,011
281	20	0,982	±	0,010	1,000	±	0,002	0,964	±	0,021	0,981	± 0,011
290	34	0,965	±	0,013	1,000	±	0,001	0,930	±	0,027	0,964	± 0,015
292	92	0,973	±	0,013	0,999	±	0,003	0,946	±	0,027	0,972	± 0,014
293	17	0,963	±	0,014	1,000	±	0,002	0,926	±	0,028	0,961	± 0,015
300	1	0,889	±	0,074	0,969	±	0,171	0,779	±	0,148	0,862	± 0,156
300	7	0,982	±	0,011	1,000	±	0,000	0,964	±	0,021	0,982	± 0,011
300	13	0,860	±	0,024	0,997	±	0,007	0,722	±	0,048	0,836	± 0,033
300	14	0,814	±	0,054	0,976	±	0,140	0,631	±	0,109	0,765	± 0,119
300	18	0,848	±	0,042	0,997	±	0,007	0,698	±	0,083	0,818	± 0,063
300	22	0,896	±	0,023	0,998	±	0,006	0,794	±	0,046	0,884	± 0,029
300	23	0,924	±	0,032	0,999	±	0,003	0,848	±	0,063	0,916	± 0,041
300	26	0,917	±	0,033	0,999	±	0,004	0,834	±	0,066	0,907	± 0,046
300	29	0,762	±	0,057	0,976	±	0,140	0,527	±	0,114	0,679	± 0,130
300	30	0,804	±	0,111	0,887	±	0,314	0,610	±	0,224	0,722	± 0,259
300	32	0,860	±	0,025	0,997	±	0,007	0,721	±	0,049	0,836	± 0,033
300	38	0,922	±	0,030	0,998	±	0,005	0,846	±	0,059	0,914	± 0,036
300	40	0,632	±	0,102	0,754	±	0,426	0,266	±	0,206	0,375	± 0,276
300	41	0,980	±	0,013	1,000	±	0,001	0,960	±	0,025	0,979	± 0,013

300	42	0,963	±	0,013	1,000	±	0,002	0,927	±	0,027	0,962	± 0,015
300	50	0,948	±	0,017	1,000	±	0,001	0,895	±	0,034	0,944	± 0,019
300	51	0,965	±	0,014	0,999	±	0,003	0,930	±	0,028	0,963	± 0,015
300	55	0,978	±	0,013	1,000	±	0,003	0,957	±	0,026	0,978	± 0,014
300	56	0,928	±	0,021	0,999	±	0,004	0,856	±	0,042	0,922	± 0,025
300	57	0,984	±	0,010	1,000	±	0,002	0,968	±	0,021	0,983	± 0,011
300	58	0,923	±	0,024	0,998	±	0,004	0,847	±	0,047	0,916	± 0,028
300	63	0,953	±	0,015	0,999	±	0,003	0,908	±	0,031	0,951	± 0,017
300	71	0,934	±	0,020	0,998	±	0,004	0,870	±	0,039	0,929	± 0,023
300	72	0,971	±	0,014	1,000	±	0,000	0,942	±	0,027	0,970	± 0,015
300	73	0,995	±	0,007	1,000	±	0,001	0,990	±	0,014	0,995	± 0,007
300	74	0,986	±	0,009	1,000	±	0,001	0,971	±	0,018	0,985	± 0,010
300	75	0,894	±	0,030	0,998	±	0,005	0,789	±	0,061	0,880	± 0,040
300	88	0,965	±	0,014	1,000	±	0,002	0,930	±	0,028	0,963	± 0,015
300	89	0,947	±	0,023	0,999	±	0,002	0,895	±	0,047	0,944	± 0,028
300	93	0,977	±	0,012	0,999	±	0,003	0,955	±	0,024	0,977	± 0,012
300	95	0,919	±	0,021	0,998	±	0,005	0,839	±	0,041	0,911	± 0,025
300	96	0,988	±	0,009	1,000	±	0,000	0,976	±	0,017	0,988	± 0,009
300	100	0,988	±	0,009	1,000	±	0,000	0,977	±	0,018	0,988	± 0,009
300	101	0,935	±	0,019	0,999	±	0,004	0,870	±	0,038	0,930	± 0,022
300	103	0,730	±	0,144	0,846	±	0,357	0,462	±	0,289	0,563	± 0,346
300	106	0,994	±	0,006	1,000	±	0,002	0,987	±	0,012	0,993	± 0,006
300	108	0,988	±	0,009	1,000	±	0,002	0,976	±	0,017	0,988	± 0,009

Tabela A.4 - NNS - Mapa de calor representando os valores das métricas de análise e os desvios padrão, considerando todos os usuários (ordenado com base em NC_u).

NC_u	u	$Acu \pm \sigma$	$Pre \pm \sigma$	$Sen \pm \sigma$	$F1s \pm \sigma$
3	104	0,645 ± 0,228	0,290 ± 0,456	0,290 ± 0,456	0,290 ± 0,456
4	69	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	12	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
5	67	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
6	53	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
7	81	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,557 ± 0,119	0,225 ± 0,417	0,117 ± 0,239	0,148 ± 0,286
11	66	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
12	48	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
12	78	0,708 ± 0,172	0,720 ± 0,451	0,415 ± 0,345	0,501 ± 0,363
12	91	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
13	76	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
16	64	0,664 ± 0,237	0,330 ± 0,473	0,330 ± 0,473	0,330 ± 0,473
17	43	0,855 ± 0,228	0,710 ± 0,456	0,710 ± 0,456	0,710 ± 0,456
18	45	0,755 ± 0,251	0,510 ± 0,502	0,510 ± 0,502	0,510 ± 0,502
25	107	0,771 ± 0,180	0,849 ± 0,359	0,544 ± 0,358	0,625 ± 0,351
33	44	0,810 ± 0,124	0,966 ± 0,172	0,623 ± 0,247	0,734 ± 0,223
35	10	0,579 ± 0,082	0,670 ± 0,473	0,159 ± 0,161	0,244 ± 0,222
36	35	0,561 ± 0,087	0,458 ± 0,493	0,127 ± 0,174	0,188 ± 0,240
37	8	0,997 ± 0,012	0,996 ± 0,018	0,998 ± 0,013	0,997 ± 0,011
38	3	0,832 ± 0,135	0,968 ± 0,111	0,679 ± 0,264	0,770 ± 0,217
39	47	0,929 ± 0,090	0,999 ± 0,009	0,858 ± 0,180	0,911 ± 0,125
42	98	0,518 ± 0,033	0,300 ± 0,461	0,037 ± 0,063	0,065 ± 0,108
43	28	0,762 ± 0,115	0,970 ± 0,171	0,525 ± 0,230	0,655 ± 0,227
43	61	0,803 ± 0,103	0,989 ± 0,041	0,612 ± 0,201	0,736 ± 0,173
46	94	0,540 ± 0,037	0,647 ± 0,478	0,080 ± 0,075	0,139 ± 0,125
49	99	0,683 ± 0,090	0,958 ± 0,197	0,367 ± 0,180	0,510 ± 0,207
53	59	0,515 ± 0,026	0,310 ± 0,465	0,031 ± 0,053	0,055 ± 0,092
56	105	0,722 ± 0,104	0,978 ± 0,109	0,449 ± 0,205	0,591 ± 0,198
65	16	0,612 ± 0,061	0,951 ± 0,182	0,227 ± 0,119	0,354 ± 0,158
82	36	0,785 ± 0,078	1,000 ± 0,000	0,570 ± 0,155	0,714 ± 0,132
82	79	0,594 ± 0,074	0,840 ± 0,368	0,188 ± 0,148	0,292 ± 0,206
87	2	0,731 ± 0,080	0,982 ± 0,102	0,466 ± 0,162	0,617 ± 0,159
89	37	0,520 ± 0,024	0,520 ± 0,502	0,039 ± 0,049	0,072 ± 0,086
97	39	0,660 ± 0,061	0,993 ± 0,040	0,322 ± 0,120	0,474 ± 0,144
102	83	0,577 ± 0,048	0,970 ± 0,171	0,155 ± 0,095	0,256 ± 0,138
103	27	0,955 ± 0,043	0,999 ± 0,006	0,911 ± 0,085	0,951 ± 0,049
104	77	0,862 ± 0,070	0,991 ± 0,019	0,731 ± 0,139	0,833 ± 0,099
109	62	0,865 ± 0,061	0,985 ± 0,024	0,742 ± 0,123	0,840 ± 0,083
117	87	0,544 ± 0,034	0,892 ± 0,266	0,092 ± 0,066	0,161 ± 0,107

120	24	0,726	±	0,063	0,980	±	0,033	0,461	±	0,124	0,618	± 0,121
121	54	0,783	±	0,065	0,996	±	0,014	0,569	±	0,129	0,716	± 0,105
123	86	0,645	±	0,054	0,997	±	0,013	0,291	±	0,109	0,440	± 0,133
127	4	0,661	±	0,052	0,984	±	0,034	0,328	±	0,104	0,483	± 0,119
128	31	0,906	±	0,053	0,989	±	0,017	0,821	±	0,106	0,893	± 0,068
136	6	0,656	±	0,054	0,989	±	0,035	0,316	±	0,108	0,469	± 0,124
143	25	0,559	±	0,042	0,920	±	0,273	0,118	±	0,085	0,202	± 0,131
145	33	0,985	±	0,020	0,993	±	0,012	0,978	±	0,038	0,985	± 0,021
147	85	0,800	±	0,059	0,994	±	0,014	0,603	±	0,118	0,744	± 0,092
148	68	0,770	±	0,059	0,982	±	0,026	0,549	±	0,118	0,697	± 0,102
162	97	0,962	±	0,030	0,987	±	0,016	0,937	±	0,060	0,960	± 0,033
170	19	0,970	±	0,024	0,996	±	0,009	0,943	±	0,049	0,968	± 0,026
173	102	0,685	±	0,052	0,995	±	0,015	0,372	±	0,103	0,533	± 0,111
175	21	0,862	±	0,052	0,983	±	0,021	0,738	±	0,104	0,839	± 0,070
179	15	0,651	±	0,041	0,995	±	0,016	0,304	±	0,083	0,460	± 0,097
185	5	0,790	±	0,055	0,994	±	0,014	0,584	±	0,110	0,729	± 0,090
186	80	0,604	±	0,037	1,000	±	0,000	0,208	±	0,074	0,338	± 0,103
196	70	0,633	±	0,034	0,998	±	0,011	0,267	±	0,069	0,416	± 0,086
200	82	0,538	±	0,024	0,980	±	0,141	0,076	±	0,049	0,138	± 0,083
207	65	0,685	±	0,051	0,963	±	0,039	0,384	±	0,103	0,541	± 0,108
212	90	0,721	±	0,047	0,980	±	0,028	0,452	±	0,094	0,613	± 0,092
214	11	0,745	±	0,050	0,995	±	0,013	0,493	±	0,100	0,653	± 0,091
215	60	0,927	±	0,035	0,980	±	0,016	0,873	±	0,069	0,922	± 0,041
248	84	0,631	±	0,038	0,999	±	0,007	0,261	±	0,075	0,409	± 0,095
266	49	0,898	±	0,037	0,992	±	0,011	0,802	±	0,074	0,885	± 0,047
270	9	0,605	±	0,027	1,000	±	0,000	0,211	±	0,055	0,345	± 0,076
280	46	0,871	±	0,037	1,000	±	0,000	0,743	±	0,074	0,850	± 0,050
281	20	0,916	±	0,034	0,999	±	0,003	0,833	±	0,069	0,907	± 0,041
290	34	0,860	±	0,031	0,993	±	0,011	0,725	±	0,062	0,837	± 0,042
292	92	0,911	±	0,034	0,991	±	0,012	0,830	±	0,067	0,902	± 0,041
293	17	0,817	±	0,042	0,998	±	0,005	0,634	±	0,084	0,773	± 0,064
300	1	0,920	±	0,026	0,992	±	0,010	0,847	±	0,052	0,913	± 0,030
300	7	0,938	±	0,025	0,996	±	0,007	0,880	±	0,049	0,934	± 0,029
300	13	0,922	±	0,028	0,991	±	0,010	0,852	±	0,055	0,915	± 0,033
300	14	0,774	±	0,040	0,982	±	0,018	0,559	±	0,082	0,709	± 0,068
300	18	0,722	±	0,039	0,987	±	0,020	0,449	±	0,077	0,614	± 0,075
300	22	0,686	±	0,040	0,999	±	0,006	0,373	±	0,081	0,538	± 0,084
300	23	0,605	±	0,034	0,998	±	0,011	0,211	±	0,069	0,343	± 0,092
300	26	0,863	±	0,034	0,995	±	0,008	0,728	±	0,069	0,839	± 0,047
300	29	0,642	±	0,034	0,992	±	0,017	0,287	±	0,069	0,441	± 0,082
300	30	0,807	±	0,041	0,979	±	0,020	0,627	±	0,079	0,762	± 0,061
300	32	0,767	±	0,038	0,994	±	0,010	0,538	±	0,076	0,695	± 0,066
300	38	0,956	±	0,021	0,995	±	0,008	0,916	±	0,044	0,953	± 0,024
300	40	0,701	±	0,041	0,993	±	0,015	0,404	±	0,082	0,570	± 0,081
300	41	0,941	±	0,025	0,998	±	0,005	0,883	±	0,049	0,936	± 0,029

300	42	0,931	±	0,030	0,998	±	0,005	0,865	±	0,060	0,925	±	0,035
300	50	0,840	±	0,046	0,986	±	0,015	0,689	±	0,093	0,808	±	0,068
300	51	0,953	±	0,021	0,993	±	0,009	0,912	±	0,043	0,950	±	0,024
300	55	0,910	±	0,026	0,998	±	0,005	0,821	±	0,053	0,900	±	0,032
300	56	0,828	±	0,041	0,998	±	0,006	0,657	±	0,082	0,790	±	0,061
300	57	0,812	±	0,037	1,000	±	0,000	0,625	±	0,074	0,767	±	0,058
300	58	0,672	±	0,039	0,999	±	0,004	0,345	±	0,078	0,507	±	0,087
300	63	0,960	±	0,026	0,995	±	0,007	0,924	±	0,050	0,957	±	0,028
300	71	0,973	±	0,018	0,997	±	0,006	0,948	±	0,036	0,972	±	0,019
300	72	0,689	±	0,037	0,996	±	0,010	0,380	±	0,075	0,546	±	0,078
300	73	0,900	±	0,034	1,000	±	0,000	0,799	±	0,068	0,887	±	0,043
300	74	0,783	±	0,039	0,998	±	0,007	0,567	±	0,079	0,720	±	0,065
300	75	0,842	±	0,041	0,998	±	0,005	0,685	±	0,081	0,810	±	0,057
300	88	0,991	±	0,010	0,995	±	0,008	0,986	±	0,019	0,991	±	0,010
300	89	0,677	±	0,039	1,000	±	0,000	0,353	±	0,079	0,517	±	0,085
300	93	0,943	±	0,024	0,996	±	0,007	0,889	±	0,049	0,939	±	0,028
300	95	0,893	±	0,036	0,995	±	0,009	0,790	±	0,070	0,879	±	0,045
300	96	0,948	±	0,030	1,000	±	0,000	0,896	±	0,060	0,944	±	0,034
300	100	0,936	±	0,027	1,000	±	0,000	0,871	±	0,054	0,930	±	0,031
300	101	0,843	±	0,046	0,999	±	0,006	0,686	±	0,091	0,810	±	0,065
300	103	0,808	±	0,037	0,998	±	0,006	0,617	±	0,073	0,760	±	0,056
300	106	0,896	±	0,031	0,999	±	0,005	0,792	±	0,062	0,882	±	0,040
300	108	0,920	±	0,030	0,999	±	0,003	0,841	±	0,059	0,912	±	0,036

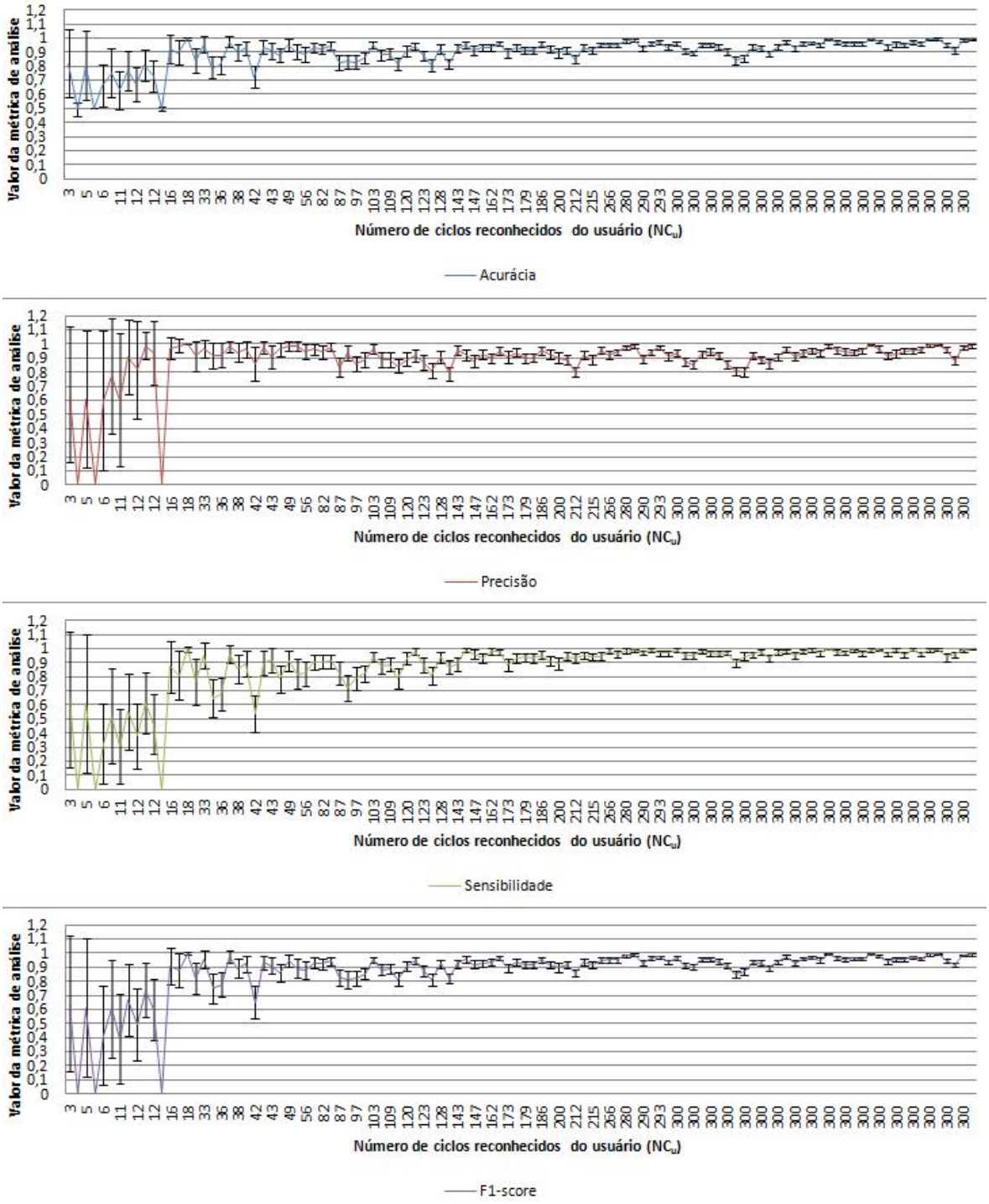

Figura A.1 - SVM - Valores das métricas de análise e dos desvios padrão por usuário

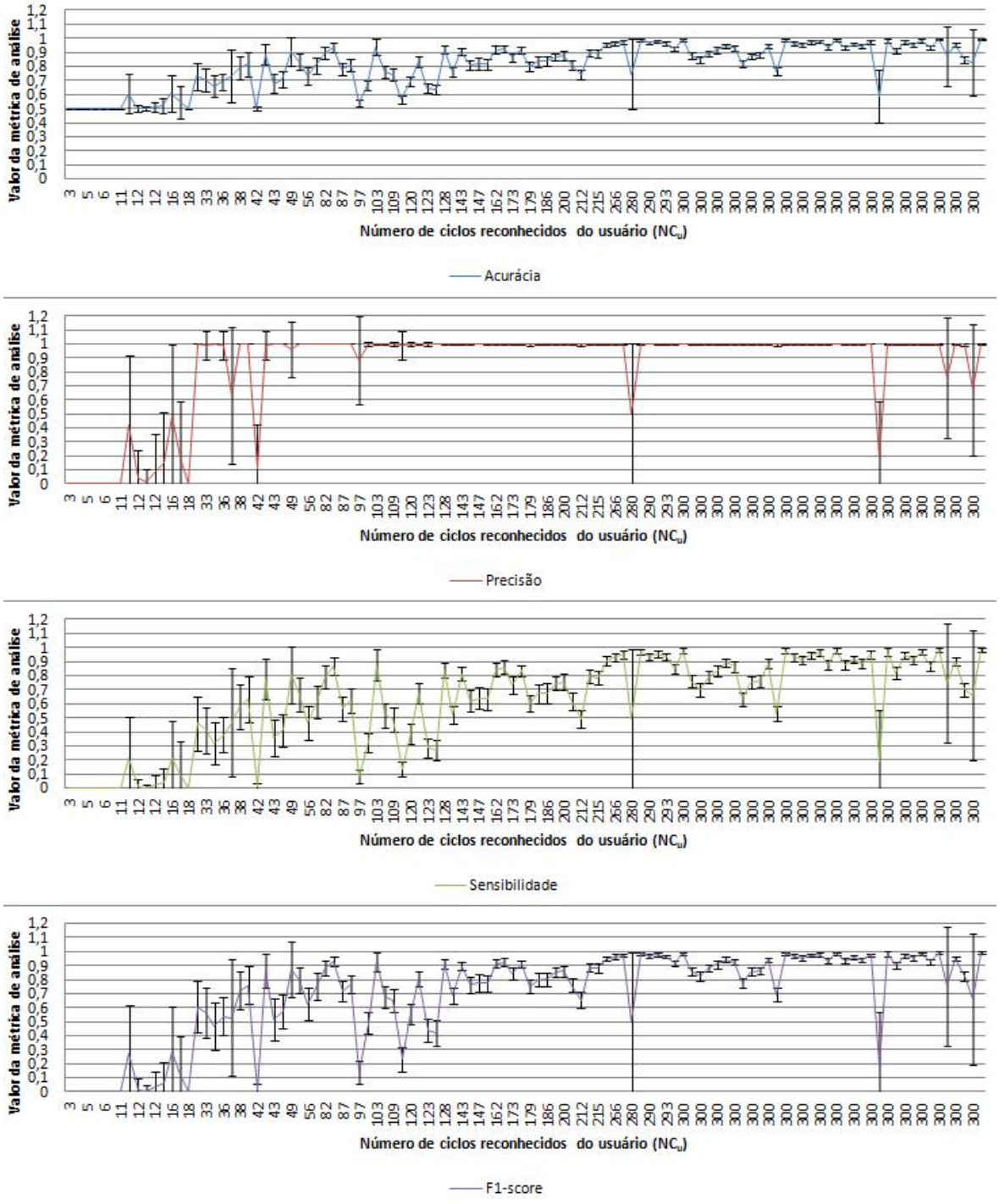

Figura A.2 - *AdaBoost* - Valores das métricas de análise e dos desvios padrão por usuário

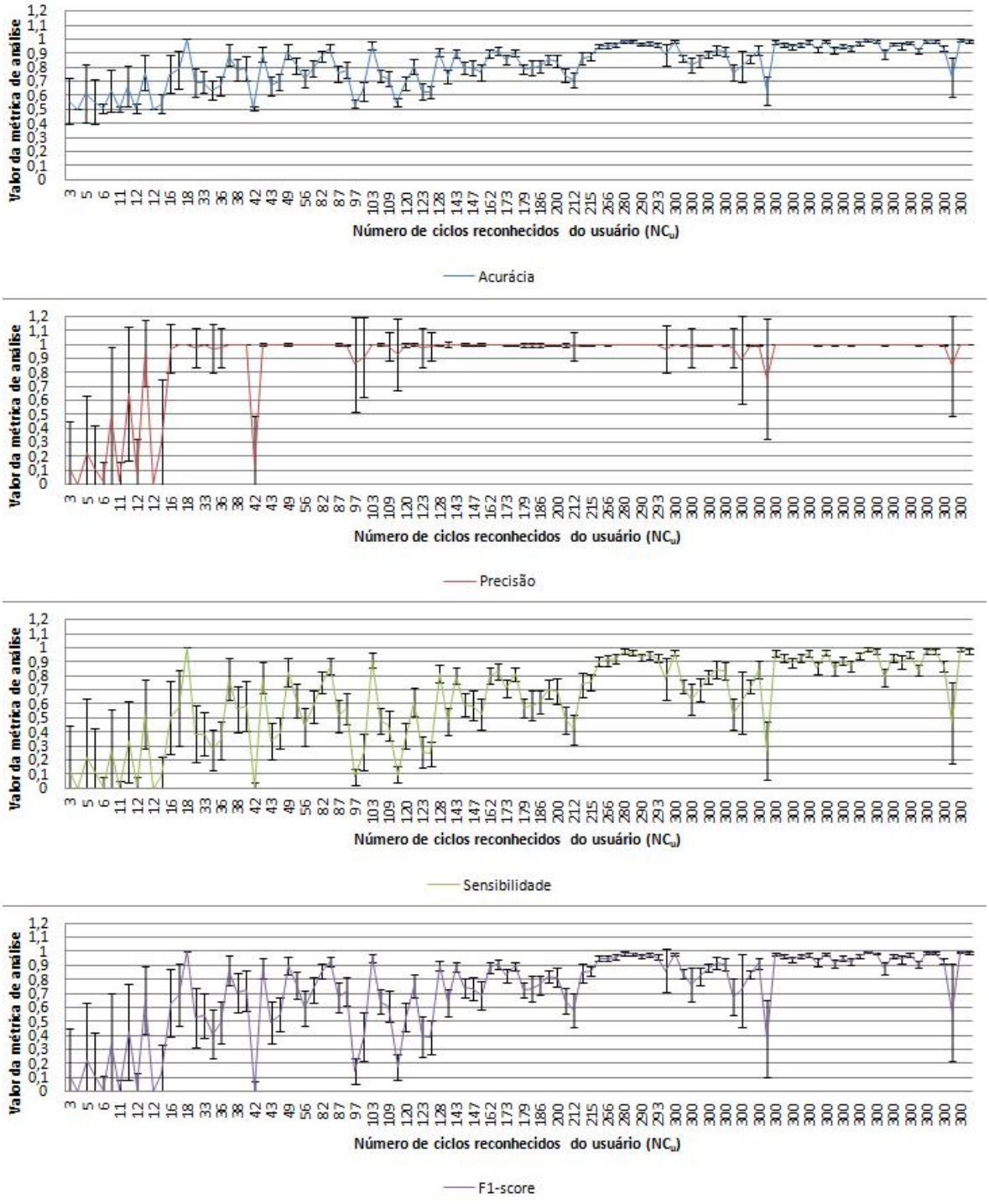

Figura A.3 - *Robust Boost* - Valores das métricas de análise e dos desvios padrão por usuário

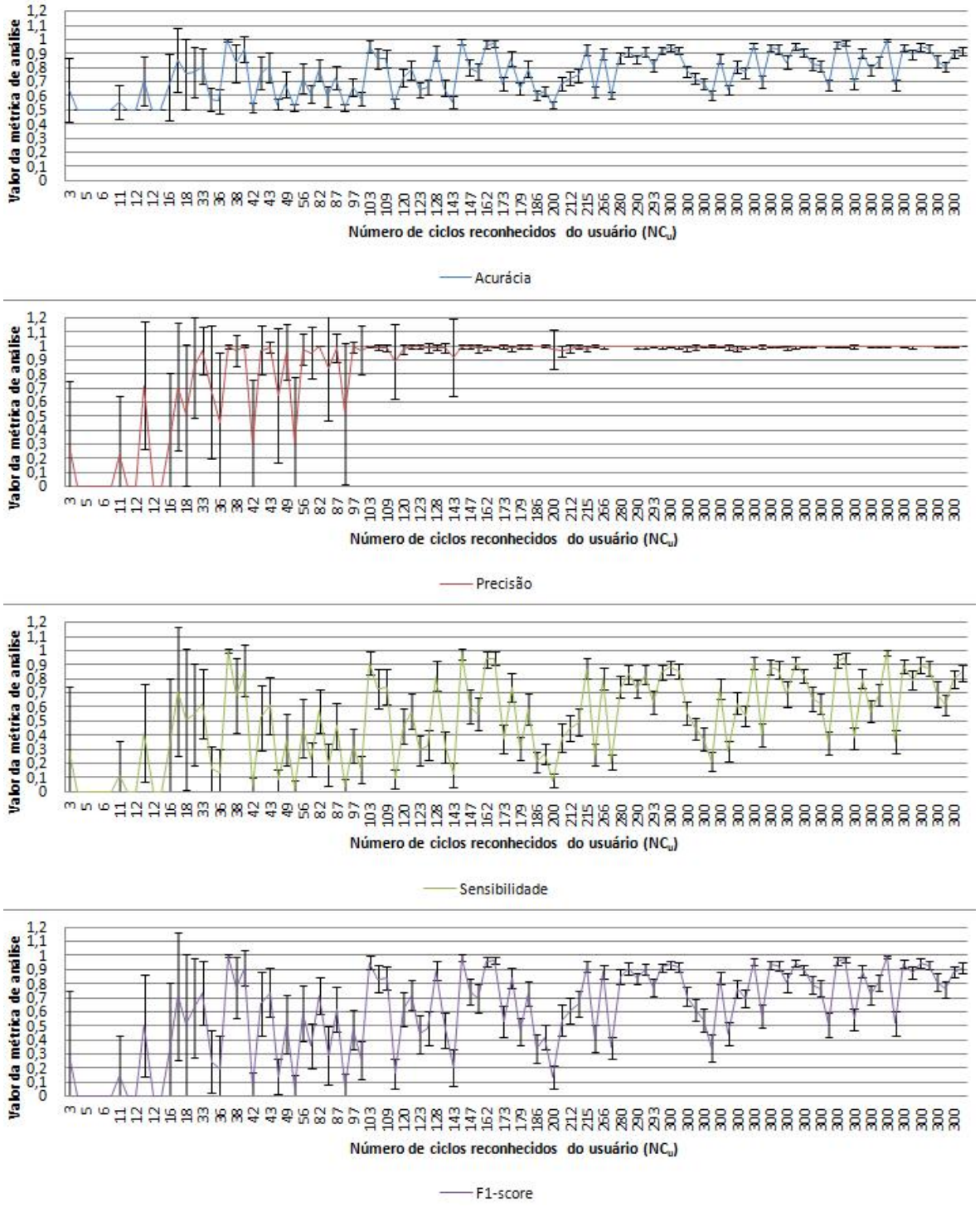

Figura A.4 - *NNS* - Valores das métricas de análise e dos desvios padrão por usuário

Tabela A.5 - SVM - Mapa de calor das médias dos tempos de treinamento ($\mu(\Delta t_{Tr})$) e de classificação ($\mu(\Delta t_{Cl})$), em segundos, por usuário (ordenado com base em NC_u).

u	NC_u	$\mu(\Delta t_{Tr})$	$\mu(\Delta t_{Cl})$	u	NC_u	$\mu(\Delta t_{Tr})$	$\mu(\Delta t_{Cl})$	u	NC_u	$\mu(\Delta t_{Tr})$	$\mu(\Delta t_{Cl})$
104	3	0,872	0,002	27	103	3,153	0,003	7	300	4,122	0,014
69	4	1,367	0,002	77	104	6,626	0,008	13	300	8,882	0,025
12	5	0,531	0,002	62	109	7,15	0,009	14	300	8,054	0,023
67	5	0,744	0,002	87	117	9,744	0,012	18	300	5,461	0,015
53	6	1,12	0,002	24	120	7,144	0,01	22	300	5,184	0,015
81	7	2,779	0,003	54	121	5,214	0,006	23	300	5,834	0,016
52	11	3,758	0,003	86	123	8,115	0,013	26	300	9,04	0,022
66	11	1,659	0,003	4	127	10,691	0,015	29	300	12,221	0,032
48	12	3,525	0,003	31	128	5,666	0,008	30	300	11,1	0,029
78	12	0,944	0,003	6	136	12,582	0,018	32	300	5,234	0,015
91	12	1,027	0,002	25	143	4,861	0,008	38	300	7,519	0,02
76	13	0,953	0,002	33	145	5,672	0,01	40	300	10,158	0,026
64	16	2,358	0,004	85	147	7,524	0,011	41	300	7,245	0,02
43	17	1,713	0,003	68	148	5,261	0,01	42	300	3,336	0,01
45	18	0,618	0,002	97	162	5,025	0,009	50	300	5,731	0,016
107	25	5,69	0,004	19	170	3,968	0,008	51	300	4,627	0,013
44	33	3,529	0,003	102	173	8,198	0,013	55	300	4,745	0,014
10	35	4,708	0,004	21	175	5,308	0,01	56	300	5,51	0,016
35	36	4,631	0,004	15	179	6,576	0,014	57	300	1,14	0,004
8	37	1,743	0,004	5	185	7,354	0,014	58	300	3,371	0,009
3	38	3,816	0,003	80	186	3,715	0,007	63	300	4,667	0,015
47	39	2,708	0,004	70	196	6,871	0,014	71	300	4,126	0,014
98	42	6,191	0,004	82	200	7,474	0,014	72	300	3,689	0,012
28	43	2,553	0,004	65	207	7,366	0,016	73	300	1,562	0,005
61	43	6,59	0,004	90	212	12,743	0,023	74	300	2,971	0,007
94	46	2,763	0,003	11	214	6,289	0,013	75	300	5,572	0,016
99	49	1,53	0,003	60	215	7,496	0,015	88	300	5,23	0,015
59	53	2,359	0,004	84	248	5,859	0,013	89	300	4,668	0,013
105	56	5,919	0,005	49	266	5,172	0,013	93	300	3,616	0,012
16	65	4,232	0,004	9	270	3,973	0,013	95	300	4,55	0,013
36	82	3,279	0,004	46	280	2,753	0,007	96	300	1,516	0,005
79	82	2,514	0,004	20	281	2,039	0,007	100	300	1,396	0,004
2	87	8,531	0,01	34	290	5,369	0,015	101	300	5,185	0,014
37	89	5,904	0,008	92	292	4,417	0,014	103	300	7,626	0,019
39	97	8,592	0,01	17	293	4,553	0,013	106	300	3,013	0,01
83	102	7,701	0,011	1	300	6,619	0,019	108	300	1,626	0,005

Tabela A.6 - *Ada* - Mapa de calor das médias dos tempos de treinamento ($\boldsymbol{\mu(\Delta t_{Tr})}$) e de classificação ($\boldsymbol{\mu(\Delta t_{Cl})}$), em segundos, por usuário (ordenado com base em $\boldsymbol{NC_u}$).

$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$
104	3	0,057	0,001	27	103	0,923	0,022	7	300	3,924	0,076
69	4	0,06	0,002	77	104	4,005	0,069	13	300	4,139	0,076
12	5	0,053	0,001	62	109	4,012	0,072	14	300	4,098	0,075
67	5	0,056	0,001	87	117	4,533	0,078	18	300	4,119	0,077
53	6	0,055	0,001	24	120	4,133	0,074	22	300	4,121	0,076
81	7	0,064	0,001	54	121	3,913	0,07	23	300	4,144	0,077
52	11	3,288	0,065	86	123	4,232	0,077	26	300	4,195	0,079
66	11	0,183	0,005	4	127	4,145	0,072	29	300	4,296	0,077
48	12	3,435	0,064	31	128	4,017	0,073	30	300	4,135	0,076
78	12	0,055	0,001	6	136	4,159	0,072	32	300	4,087	0,076
91	12	3,441	0,069	25	143	4,101	0,074	38	300	4,02	0,075
76	13	0,833	0,016	33	145	3,988	0,073	40	300	4,307	0,077
64	16	0,141	0,005	85	147	3,822	0,072	41	300	3,953	0,076
43	17	0,086	0,002	68	148	4,182	0,074	42	300	3,979	0,076
45	18	0,052	0,002	97	162	3,773	0,073	50	300	4,177	0,075
107	25	3,6	0,071	19	170	3,898	0,073	51	300	3,911	0,074
44	33	3,733	0,07	102	173	4,12	0,076	55	300	3,772	0,076
10	35	3,923	0,071	21	175	4,454	0,089	56	300	4,005	0,077
35	36	3,971	0,07	15	179	4,017	0,072	57	300	2,762	0,058
8	37	0,153	0,005	5	185	4	0,073	58	300	4,008	0,076
3	38	3,74	0,071	80	186	4,32	0,08	63	300	4,161	0,08
47	39	3,43	0,066	70	196	4,158	0,072	71	300	3,928	0,074
98	42	4,054	0,074	82	200	4,506	0,086	72	300	3,882	0,073
28	43	1,585	0,034	65	207	4,162	0,073	73	300	0,067	0,002
61	43	3,787	0,071	90	212	4,342	0,084	74	300	2,791	0,055
94	46	3,819	0,07	11	214	4,065	0,074	75	300	4,232	0,073
99	49	2,121	0,044	60	215	4,111	0,074	88	300	4,079	0,088
59	53	4,034	0,071	84	248	4,041	0,076	89	300	4,325	0,088
105	56	3,94	0,073	49	266	3,951	0,073	93	300	3,891	0,081
16	65	4,318	0,083	9	270	4,073	0,074	95	300	4,026	0,076
36	82	3,823	0,073	46	280	0,128	0,003	96	300	3,489	0,074
79	82	4,06	0,078	20	281	3,965	0,077	100	300	0,346	0,009
2	87	3,999	0,072	34	290	4,007	0,076	101	300	4,127	0,077
37	89	4,115	0,072	92	292	3,831	0,08	103	300	4,092	0,077
39	97	4,169	0,072	17	293	4,104	0,084	106	300	0,25	0,006
83	102	4,345	0,082	1	300	4,202	0,077	108	300	2,939	0,067

Tabela A.7 - *Rob* - Mapa de calor das médias dos tempos de treinamento ($\boldsymbol{\mu(\Delta t_{Tr})}$) e de classificação ($\boldsymbol{\mu(\Delta t_{Cl})}$), em segundos, por usuário (ordenado com base em $\boldsymbol{NC_u}$).

$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	$\boldsymbol{u}$	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$
104	3	3,254	0,052	27	103	2,783	0,038	7	300	3,644	0,049
69	4	3,803	0,049	77	104	2,807	0,031	13	300	3,717	0,041
12	5	3,492	0,05	62	109	2,92	0,035	14	300	3,022	0,034
67	5	3,528	0,049	87	117	1,972	0,02	18	300	2,485	0,028
53	6	3,603	0,049	24	120	3,461	0,038	22	300	2,501	0,03
81	7	3,788	0,059	54	121	2,636	0,03	23	300	3,866	0,045
52	11	2,841	0,039	86	123	2,31	0,025	26	300	3,986	0,045
66	11	3,458	0,046	4	127	2,17	0,023	29	300	2,451	0,027
48	12	2,96	0,04	31	128	3,72	0,044	30	300	2,867	0,031
78	12	3,36	0,054	6	136	3,185	0,033	32	300	2,359	0,028
91	12	3,007	0,041	25	143	4,012	0,046	38	300	2,99	0,037
76	13	3,574	0,043	33	145	2,788	0,033	40	300	1,987	0,022
64	16	3,331	0,046	85	147	2,747	0,032	41	300	3,817	0,052
43	17	3,4	0,049	68	148	2,425	0,029	42	300	3,909	0,048
45	18	3,413	0,054	97	162	4,22	0,05	50	300	4,662	0,055
107	25	3,011	0,04	19	170	3,878	0,047	51	300	3,164	0,041
44	33	3,306	0,044	102	173	3,887	0,044	55	300	3,66	0,052
10	35	3,256	0,041	21	175	4,809	0,058	56	300	3,465	0,042
35	36	3,185	0,039	15	179	4,096	0,047	57	300	3,161	0,045
8	37	3,158	0,044	5	185	2,402	0,028	58	300	3,168	0,039
3	38	3,161	0,042	80	186	3,122	0,038	63	300	4,141	0,051
47	39	3,173	0,043	70	196	3,178	0,035	71	300	3,07	0,038
98	42	2,718	0,03	82	200	3,277	0,036	72	300	3,784	0,048
28	43	3,135	0,042	65	207	1,883	0,02	73	300	3,776	0,052
61	43	3,226	0,039	90	212	2,81	0,032	74	300	2,986	0,039
94	46	2,561	0,031	11	214	3,142	0,035	75	300	2,895	0,034
99	49	3,332	0,046	60	215	3,257	0,037	88	300	4,349	0,059
59	53	3,671	0,045	84	248	4,676	0,057	89	300	5,016	0,058
105	56	3,378	0,04	49	266	3,355	0,041	93	300	3,877	0,051
16	65	4,057	0,057	9	270	3,914	0,046	95	300	2,351	0,028
36	82	3,576	0,046	46	280	3,571	0,054	96	300	3,198	0,046
79	82	3,159	0,043	20	281	3,617	0,049	100	300	3,629	0,05
2	87	3,124	0,037	34	290	4,204	0,052	101	300	3,643	0,044
37	89	3,665	0,041	92	292	3,755	0,049	103	300	1,869	0,02
39	97	1,972	0,02	17	293	4,375	0,056	106	300	3,47	0,05
83	102	2,795	0,031	1	300	2,723	0,032	108	300	3,248	0,045

Tabela A.8 - *NNS* - Mapa de calor das médias dos tempos de treinamento ($\boldsymbol{\mu(\Delta t_{Tr})}$) e de classificação ($\boldsymbol{\mu(\Delta t_{Cl})}$), em segundos, por usuário (ordenado com base em $\boldsymbol{NC_u}$).

u	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	u	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$	u	$\boldsymbol{NC_u}$	$\boldsymbol{\mu(\Delta t_{Tr})}$	$\boldsymbol{\mu(\Delta t_{Cl})}$
104	3	< 0,001	< 0,001	27	103	< 0,001	0,001	7	300	< 0,001	0,002
69	4	< 0,001	< 0,001	77	104	< 0,001	< 0,001	13	300	< 0,001	0,001
12	5	< 0,001	0,001	62	109	< 0,001	< 0,001	14	300	< 0,001	0,001
67	5	< 0,001	< 0,001	87	117	< 0,001	0,001	18	300	< 0,001	0,002
53	6	< 0,001	0,001	24	120	< 0,001	0,001	22	300	< 0,001	0,001
81	7	< 0,001	0,001	54	121	< 0,001	0,001	23	300	< 0,001	0,001
52	11	< 0,001	< 0,001	86	123	< 0,001	0,001	26	300	< 0,001	0,002
66	11	< 0,001	0,001	4	127	< 0,001	0,001	29	300	< 0,001	0,002
48	12	< 0,001	< 0,001	31	128	< 0,001	0,001	30	300	< 0,001	0,001
78	12	< 0,001	< 0,001	6	136	< 0,001	0,001	32	300	< 0,001	0,002
91	12	< 0,001	0,001	25	143	< 0,001	0,001	38	300	< 0,001	0,001
76	13	< 0,001	< 0,001	33	145	< 0,001	0,001	40	300	< 0,001	0,001
64	16	< 0,001	0,001	85	147	< 0,001	0,001	41	300	< 0,001	0,002
43	17	< 0,001	< 0,001	68	148	< 0,001	< 0,001	42	300	< 0,001	0,001
45	18	< 0,001	< 0,001	97	162	< 0,001	0,002	50	300	< 0,001	0,002
107	25	< 0,001	< 0,001	19	170	< 0,001	0,001	51	300	< 0,001	0,001
44	33	< 0,001	< 0,001	102	173	< 0,001	0,001	55	300	< 0,001	0,002
10	35	< 0,001	< 0,001	21	175	< 0,001	0,002	56	300	< 0,001	0,001
35	36	< 0,001	< 0,001	15	179	< 0,001	0,001	57	300	< 0,001	0,001
8	37	< 0,001	< 0,001	5	185	< 0,001	0,001	58	300	< 0,001	0,002
3	38	< 0,001	0,001	80	186	< 0,001	0,001	63	300	< 0,001	0,001
47	39	< 0,001	< 0,001	70	196	< 0,001	0,001	71	300	< 0,001	0,001
98	42	< 0,001	< 0,001	82	200	< 0,001	0,001	72	300	< 0,001	0,001
28	43	< 0,001	< 0,001	65	207	< 0,001	0,001	73	300	< 0,001	< 0,001
61	43	< 0,001	< 0,001	90	212	< 0,001	0,001	74	300	< 0,001	0,002
94	46	< 0,001	< 0,001	11	214	< 0,001	0,001	75	300	< 0,001	0,001
99	49	< 0,001	0,001	60	215	< 0,001	0,001	88	300	< 0,001	0,001
59	53	< 0,001	< 0,001	84	248	< 0,001	0,001	89	300	< 0,001	0,001
105	56	< 0,001	0,001	49	266	< 0,001	< 0,001	93	300	< 0,001	0,001
16	65	< 0,001	< 0,001	9	270	< 0,001	0,001	95	300	< 0,001	0,001
36	82	< 0,001	0,001	46	280	< 0,001	0,001	96	300	< 0,001	0,002
79	82	< 0,001	< 0,001	20	281	< 0,001	0,001	100	300	< 0,001	0,001
2	87	< 0,001	< 0,001	34	290	< 0,001	0,002	101	300	< 0,001	0,001
37	89	< 0,001	0,001	92	292	< 0,001	0,001	103	300	< 0,001	0,001
39	97	< 0,001	0,001	17	293	< 0,001	0,002	106	300	< 0,001	0,001
83	102	< 0,001	0,001	1	300	< 0,001	0,001	108	300	< 0,001	0,001

APÊNDICE B - ARTIGO SUBMETIDO A PERIÓDICO COM JCR

AUTHENTICATION BASED ON ELECTROCARDIOGRAPHY SIGNALS AND MACHINE LEARNING IN A MOBILE CLOUD COMPUTING SCENARIO

Silas Leite Albuquerque , Paulo Roberto L. Gondim and Cristiano Jacques Miosso

Department of Electrical Engineering, Brasília University (UnB), Brasília, Brazil

ABSTRACT

Mobile Cloud Computing (MCC) is a highly complex topic that encompasses several information security problems. The authentication area of the various entities involved has been extensively discussed in recent years and shown a wide range of possibilities. The use of inadequate authentication processes leads to several problems, which range from financial damage to users or providers of Mobile Commerce (M-Commerce) services to the death of patients who depend on Mobile Healthcare (M-Health) services.

The design of reliable authentication processes that minimize such issues involves the needs for using non-intrusive authentication techniques and continuous authentication of users by MCC service providers. In this sense, biometrics could be used for supporting these needs in various scenarios.

This research has explored some conceptual bases and presents a continuous authentication protocol for MCC environments. The protocol is part of a cyber-physical system (CPS) and is based on the monitoring of physiological information interpreted from electrocardiograms (ECG) of a database with 108 users. Machine learning techniques based on the Near Neighbor Search (NNS) and the Ensemble Adaptative Boost (or AdaBoost) were used for the classification of the heart cycles recognized in such ECGs.

The two ML techniques applied to electrocardiography were compared by a random subsampling technique that considers four analysis metrics, namely accuracy, precision, sensitivity and F1-score. The experimental results showed the better performance of AdaBoost regarding accuracy (92,0%), sensitivity (84.0%) and F1-score (89.3%). On the other hand, NNS provided better accuracy (99.5%).

1. INTRODUCTION

Mobile Cloud Computing (MCC) has arisen from the combination of the flexibility of mobile computing and the storage and processing capabilities of cloud computing [1]. Most processing and storage of data from mobile devices are transferred to centralized computing platforms located in the cloud, thus enabling such devices of lower computational capacities to run more complex applications and access new resources and services [2], as long as they are connected to the cloud through the various available technologies (e.g. local wireless networks (WiFi) and cellular networks (4G, 5G).

Figure 1 shows a typical and simplified architecture of both MCC and the services that can be provided. It displays mobile users with their equipment accessing computing cloud services (Infrastructure as a service - IaaS, Platform as a service – PaaS and Software as a service - SaaS) through conventional wireless Internet access connections [3].

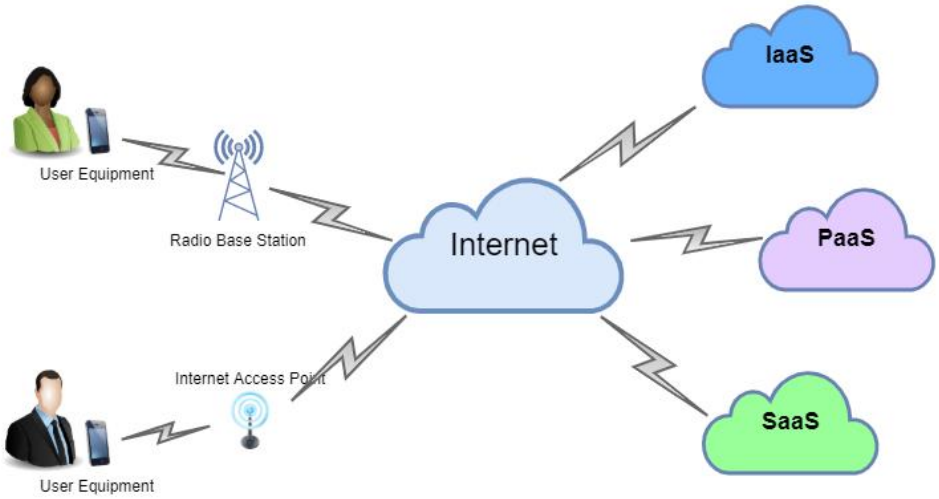

Figure 1 - Typical architecture of MCC services

Such a style of use of computer environments has evolved in recent years greatly due to the significant expansion of the smart phone market, which has increased the number of mobile computing users and, consequently, required qualitative and quantitative improvement in the infrastructures focused on this segment.

Despite the benefits of cloud mobile computing, new problems have arisen. The area of information security, for example, shows a number of loopholes that did not exist in other more traditional architectures [4]. Because cloud computing is based on the remote use of information for either its simple storage, or its processing, several fundamental pillars of information security (e.g., confidentiality, integrity, availability and authenticity) have been threatened [5].

Due to the considerable increase in the number of elements between the user interface and the place where the information will be stored and/or processed in the cloud, in comparison to isolated architectures, the authentication of the integral parts of an MCC architecture has drawn the attention of academic and commercial circles [6].

[7], [8], [9], [10] and [11] analyzed and/or proposed authentication solutions for MCC environments, which indicates the existence of many widely used authentication methods and protocols globally recognized by researchers and even standardized by some regulatory bodies. Among such methods, some biometric techniques have proved interesting alternatives, not only because they depend exclusively on intrinsic human aspects to be authenticated, but also because they enable the continuity of authentication (the user can remain authenticated throughout the session [12] in a transparent or non-intrusive manner), besides the authentication at an early stage of a session or transaction (e.g., access of user´s equipment to the MCC infrastructure).

Electrocardiography-based biometrics is one of such options. Its characteristics are unique, can be measured, occur with no individual´s voluntary intervention, or intervention by an individual from whom they are extracted, and are hardly falsified [13], [14], [15]. Therefore, the method can be used in activities of authentication of people by automated information systems.

Although electrocardiography is a good option for authentication, the misinterpretation of data from ECGs can lead to distortions. Machine learning (ML) [16] has shown a flexible and robust alternative for minimizing such a problem, since it provides good solutions to complex classification issues and effectiveness and efficiency for authentication.

ML techniques include ensembles (e.g., AdaBoost, Robust Boost, Random Forest), which represent a strategy in which multiple

simple learners are trained and then used in combination [17]. The ensembles can provide better performance, in terms of precision, sensitivity, and accuracy, than that of the best simple classifier from the set used individually, and even, in many cases, than that of different, more complex classifiers trained under other strategies [31]. For this reason, ensembles have been used in several tasks involving signal and data classification, such as voice recognition [32], adaptive face recognition [33], and brain degenerative diseases diagnosis [34].

This article proposes a non-intrusive and continuous authentication protocol for MCC environments, and reports on its evaluation regarding effectiveness and efficiency, as well as a comparative analysis with other schemes from the literature. Machine learning techniques that provide levels of appropriate accuracy, precision and sensitivity are addressed from characteristics extracted by temporal analysis and amplitude of the ECG signal.

Our protocol is inserted in a Cyber-physical System (CPS), since it integrates physical and software components towards monitoring vital signs and using them as a method of authentication of individuals by systems. The article also addresses a proposal for future work recently presented in [18] regarding the phasing for authentication, but with no implementation.

On the other hand, AdaBoost is an ensemble-type machine learning technique that relies on a boosting algorithm for classification after a supervised training stage, and which can provide comparatively high performance even from relatively few training examples. It has been used in several real-world classification tasks, including the classification of physiological signals and even of ECG abnormalities [35]. However, to the best of our knowledge, no previous research evaluated its use for user authentication based on healthy ECG signals.

In this manuscript, AdaBoost was employed and evaluated in comparison with the technique based on NNS (Nearest Neighbor Search) and proposed in [13]. Better levels of accuracy, sensitivity and f1-score were obtained. Our motivation for its use was based on its previous mentioned success in the classification of physiological signals, under other contexts.

The proposed protocol is based on fiducial points, which are specific and relevant points of the cardiac signal, here used for the obtention of temporal and amplitude-based inter-point distances, to be adopted as a part of the user identification process. The detection of such fiducial points has been recognized as of great importance for the success of the ECG-based authentication process ([19]). Moreover, the normalization of temporal inter-point distances of ECG signals should be investigated, in terms of the possible results improvement in the classification process based on ML techniques.

Below are the main contributions of this research:

- Implementation of a new fiducial point detection algorithm based on a pre-defined pattern of the cardiac cycle;
- Proposal and evaluation of a new way of normalizing the temporal characteristics of the cardiac signal towards good quality features for ML processes;
- Use of AdaBoost for the classification of electrocardiography signals for user authentication, with better results than NNS in most part of metrics;
- Application of the random subsampling technique for the classification processes and obtaining of their analysis metrics (accuracy, precision, sensitivity and f1-score); and
- Proposal of an authentication protocol based on analysis of ECG signals and ML techniques.

The remainder of the article is organized as follows: Section 2 briefly describes and compares some articles on the themes explored in this research; Section 3 presents the authentication protocol with its several phases, methods and algorithms used; Section 4 is devoted to the evaluation of the

protocol focusing on aspects related to electrocardiography and machine learning techniques; finally, Section 5 provides the conclusions and suggests some future work.

2. RELATED WORK

This section addresses some articles related to authentication schemes, protocols or models that use electrocardiography for authentication, with possible application in MCC environments. Some studies have analyzed and / or proposed authentication and / or identification solutions with the use of electrocardiography, and the two main approaches are "fiducial" (based on fiducial points) and “non-fiducial” (based on other techniques). Fiducial points based approaches were preferably chosen since normalization algorithms can be used to compensate for changes in ECG signal due to the heart rate variability ([19]).

- D. Rezgui and Z. Lachiri [14] identified individuals based on their electrocardiograms. 21 features chosen were derived from fiducial points (distances in time and tension between points), and 10 morphological descriptors were obtained from the analysis of each cardiac cycle. The authors criticize other related studies stating they are strongly based on approaches that consider only the locations of the fiducial points, which generates attributes that represent only some specificities of the electrocardiograms. Therefore, the authors consider morphological descriptors that add analysis parameters to the process as a whole. A problem, however, concerns the automatic QRS detector, which, according to the authors' own evaluations, is not very accurate. Another aspect that can be criticized is the use of "features derived from other features", which indicates the redundant use of information and the probable unnecessary increase in the analyses conducted. The main result found in this work points to an accuracy of 99.38%.

- J. S. Arteaga-Falconi, H. A. Osman and A. E. Saddik [13] proposed a new biometric authentication algorithm based on electrocardiography and aimed at mobile computing (according to the authors, this is the first algorithm published). The approach considered for ECG-based authentication focuses on the use of 8 features derived from fiducial points (distances in time and electrical potential between points) that have no redundancies. 73 users´ records from 4 public ECG bases and 10 records captured by a sensor with 2 electrodes (each electrode is in contact with a finger of a different hand) were used. However, the authors do not clarify the way they detected the fiducial points, and the alignment phase is considered unnecessary in the algorithm, since it is carried out implicitly when the features are calculated by the subtractions indicated in the fourth phase of the process. The final result from this work indicates an accuracy of 81.82%.
- C. Camara, P. Peris-Lopez, L. Gonzalez-Manzano and J. Tapiador [15] designed a continuous authentication system based on ECG and aimed at real-time applications. According to the authors, the area of biometric systems has greatly evolved and the variety of biosignals available for these systems has increased. They also proposed using ML and Data Stream Mining (DSM) techniques for the obtaining of good results from biometric systems. The system works from a general structure divided into configuration and operation phases. Differently from other approaches (based on fiducial points), the one used for the extraction of features is considered "non-fiducial" and is based on the obtaining of features from the application of the "Walsh-Hadamard" transform. kNN machine learning technique is used for the learning process. A criticism is, however, the existence of a similarity module that discards samples. Despite focusing on samples that fit the first model created, the approach may neglect those representative of the individual. The main result found in this

work points to an accuracy of 94.79%.

- In [20], J. S. Arteaga-Falconi, H. A. Osman and A. E. Saddik proposed an evolution of the previous work adding another factor of authentication based on fingerprint to the original algorithm. The approach used in 2016 for the ML classification was based on NNS, was replaced by one that uses SVM. The same public ECG databases were used for ECG analysis. The ECGs were divided into 60-second fragments for the training of the SVM and 4-second ones for the classification tests. Despite the improvements from the SVM application, the use of the fingerprint-based technique suppressed the characteristics of non-intrusiveness and continuity observed in [13]. At the end, the authors indicate that, using this approach, they archive an accuracy of 100%.

3. PROPOSED PROTOCOL

3.1. INITIAL CONSIDERATIONS

This section introduces a protocol for a non-intrusive and continuous authentication for users of mobile cloud computing environments - MCC.

The system uses a factor for the authentication of each user: "something intrinsic to the person" - their cardiac cycles represented, in this context, by an electrocardiogram (ECG). The use of cardiac cycle data as biometric features of authentication can be considered a non-intrusive and transparent technique for the user. The technique offers security advantages over other biometric ones, such as voice recognition (which is subject to an attack based on the reproduction of a recording of the user's voice), facial recognition (which is subject to an attack based on the presentation of a captured image of the user), and fingerprint recognition (which cannot be considered transparent to the user).

The biometric features available from the use of electrocardiography can be classified by machine learning (ML) techniques. Such ML techniques include classifiers (for example, SVM, ensembles, NNS, among others). On the other hand, MCC could be used for the sake of storing and processing information related to mobile users, including ECG signals, in a CSP. Then, the possibility of employment of such ML techniques in the context of MCC should be investigated. Apart from the factor aimed at user authentication by the CSP, the provider is also authenticated by the user through a digital certificate found at the CSP, and the use of a SSL / TLS connection. Therefore, the protocol can be considered a mutual authentication one between user and provider.

3.2. THE NETWORK MODEL

Figure 2 shows the network model on which our authentication protocol should work. The main entities are:

- Cardiac sensor (HS), which captures the user's cardiac signals and provides them for the creation of an ECG representative of the individual;
- User equipment (UE), through which the user authenticates and receives the services made available by the service provider;
- Cloud Computing Service Provider (CSP), which centralizes requests and provision of cloud computing services. It must verify the authenticity of the plaintiffs (users) and authenticate itself (as previously indicated, the CSP for this model has a digital certificate linked to asymmetric cryptographic keys that enables its authentication with users);
- Continuous Authentication Center (CAC), which is a set of entities (real or virtual machines) of the computational cloud allocated for the various tasks related to the authentication processes;
- Authentication Database (ADB), which stores data to be used in authentication processes.

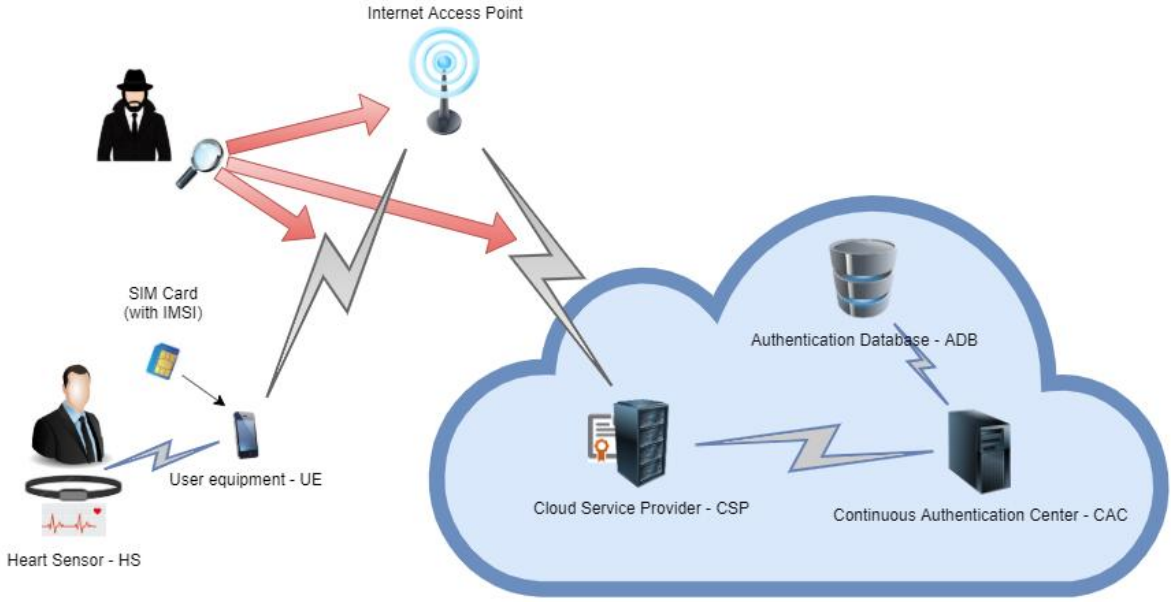

Figure 2- Network model considered

The protocol aims at an end-to-end authentication and is designed to work at the application layer, and several communication technologies could be used in the lower layers. For instance, the HS may communicate with the UE through a Bluetooth connection, the UE may communicate with the CSP through an Internet connection via WLAN or cellular networks (3G, 4G, 5G), and the internal entities of the cloud may communicate using the technologies available in the cloud itself (e.g., Ethernet connections that use different cables, optical fibers or even establish wireless communications).

For the model indicated, the communication channels considered insecure and, therefore, subjected to intervention of possible attackers or intruders, are those between the UE and the CSP, as shown in figure 2, where a potential attacker is monitoring the indicated channels. The other environments (region close to the user where the individual, HS and UE are present, and internal cloud environment formed by CSP, CAC and ADB) are considered free from threats.

3.3. PROTOCOL PHASING

The proposed authentication protocol is divided into the following three phases:

- Registration, in which the user introduces himself to the system and has some of his data collected (e.g., raw electrocardiographic signals) and stored for the following phases.
- Authentication, in which the user presents new data that, after being compared with those already stored, generate the acceptance or rejection of the user's access.
- Update, in which new data replace those stored in the registration phase.

3.3.1 Registration

In this phase, performed before the authentication protocol begins the authentication, relevant user´s data are captured and sent to the cloud to be stored in the ADB for future authentication processes. The phase must be triggered in a controlled physical environment where the CSP is secure regarding the user's identity.

A continuous capture of raw ECG data is foreseen until the minimum necessary number of cycles recognized for classifier

training has been obtained and the recording phase has been completed.

The registration phase consists of the following steps:

- The UE receives the raw ECG data extracted from the user via HS;
- Using the CSP digital certificate, the UE creates an SSL connection between itself and the CSP through which raw ECG data are sent;
- The CSP receives the data through the SSL connection and forwards it to the CAC;
- CAC verifies the existence, in the ADB, of the user being included;
- If the user is not found, the CAC processes the raw ECG data by recognizing its cycles, detecting the fiducial points and extracting the ECG features to be used in the authentication process (at this time, if the number of recognized cycles is lower than the standardized number of cycles for classifier training, the need for more raw ECG data is signalized to the entities involved - CSP, UE and HS - and they provide the complementation);
- ECG features are then used for training the classifier, which, after being trained, provides its training parameters to be stored in the ADB and used in the authentication processes;
- After all such processes have succeeded, CAC registers the classifier parameters in the ADB.

3.3.2. Authentication

This phase is performed whenever a user aims to authenticate with the CSP for requesting a service. Authentication is continuous and carried out iteratively, as long as the user wishes to remain authenticated by the CSP, and is composed of the following steps:

- The UE receives the raw ECG data extracted from the user through cardiac sensors;
- Raw ECG data are sent through the SSL connection between the UE and the CSP;
- The CSP forwards the data received to the CAC;
- The CAC processes the raw ECG data by extracting the ECG features to be used in the authentication process;
- The CAC authenticates the user through the classification of the newly obtained ECG features with the classifier already trained;
- If the authentication has succeeded, the CAC informs it to the CSP.

The continuous authentication is performed with each cardiac cycle. The following two situations deserve special attention: the protocol does not recognize any cycle and cannot use it for authentication, and the protocol refutes the authentication of some cardiac cycle.

Since errors may occur in the classification, an interruption in the authenticated session is unnecessary. For such cases, the protocol considers the following alternatives:

- If a cycle is not recognized, new attempts are made successively after the failure and, if the problem remains for a longer period of time than a predefined tolerance, the session is interrupted and its restart is requested (if the cycles are recognized again before tolerance time, the process proceeds normally).
- If any cycle is disproved in the authentication, similarly to the previous situation, new attempts are made towards reversing the problem (if it is reversed, the process proceeds normally). After a maximum tolerance time has been reached, the session is interrupted and blocked; it can be unlocked only after an update in a controlled environment (explained below)

.

3.3.3. Update

This phase is performed when a change must be made in one or more data related to the user that are stored in the ADB. It can be motivated by either a change in the cardiac pattern (caused by a surgery, for example), or the need (or desire) to retrain the classifier.

The update can occur in two ways:

- in a controlled environment, when the user is no longer able to authenticate with the CSP (change in the cardiac authentication pattern). It must be triggered, as in the registration phase, in a controlled physical environment where the CSP is safe in relation to the user's identity; and
- during an authenticated session, when an already authenticated user wishes to retrain the classifier.

Update in a controlled environment

The update in a controlled environment is triggered according to the following steps:

- The UE receives the data to be changed (raw ECG data);
- Using the CSP digital certificate, the UE creates an SSL connection between itself and the CSP through which raw ECG data are sent;
- The CSP forwards the data received to the CAC;
- CAC verifies the existence, in the ADB, of the user being updated;
- If the user is found, the CAC processes the raw ECG data by extracting the ECG features to be used in the authentication process (as in the registration, if not enough raw ECG data are available for the classifier training, more data must be provided);
- ECG features are then used for training the classifier, which, then, provides its training parameters to be stored in the ADB and used later in the authentication processes;
- The CAC updates the data record in the ADB modifying the classifier parameters;
- After a successful update, it then informs both CSP and user on the successful change.

Update during an authenticated session

As previously indicated, this update occurs when the user is already authenticated with the CSP and is performed as follows:

- The UE receives the data to be changed (raw ECG data);
- The raw ECG data are sent through the SSL connection between the UE and the CSP;
- CSP forwards the data received to the CAC;
- CAC processes the raw ECG data extracting the ECG features to be used in the authentication process (similarly to the registration and update in a controlled environment, if not enough raw ECG data are available for classifier training, more data must be provided);
- ECG features are then used for training the classifier, which then provides its training parameters to be stored in the ADB and used later in the authentication processes;
- CAC updates the data record in the ADB modifying the classifier parameters;
- After a successful update, it informs both CSP and user on the change and proceeds with the authenticated session.

3.4. SPECIFIC ASPECTS OF THE ELECTROCARDIOGRAPHY-BASED PROCESS

Below are some aspects related to electrocardiography:

- The proposal focuses on the authentication of people who do not have cardiac abnormalities and whose cardiac cycles follow a certain pattern (indicated below). This premise is mainly used for imposing conditions for the detection of fiducial points used in the extraction of features. If it is not adopted, the fiducial point detection algorithm used must follow another logic of operation, which is outside the scope of this research;
- The cardiac cycle adopted is a curve representing part of an ECG between a fiducial point LP and the subsequent fiducial point TP, i.e., a waveform containing fiducial points LP, P, Q, R, S, T and TP, which are the basis for the extraction process of ECG features (see Figure 3). The cycle must follow a pattern according to which,

from a fiducial point, the path to the next point must exhibit only one behavior, i.e., be ascending or descending (the path may contain horizontal segments);

- Although the fiducial point detection algorithm developed is based on the indicated pattern, it has some tolerance ranges that enable the recognition of a cycle even if shows small (configurable) variations in the desired behavior.

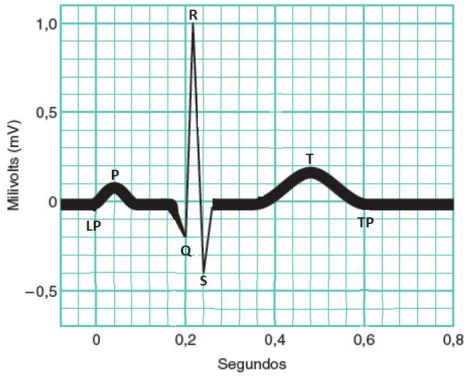

Figure 3- Typical cardiac cycle (adapted from [21])

The treatment of data derived from ECG consists of the following three main methods:

- Extraction of ECG features;
- Training of the classifier; and
- Authentication.

Below is the description of each method focused on the context of electrocardiograph.

3.4.1. ECG feature selection and extraction

In defining the ECG features on which our classification strategies are based, we adopted a set of time intervals and amplitudes that according to previous research [14], [19] are relatively consistent for the same individual, but which change from person to person. Specifically, we used the same features proposed in [13], but with a different normalization method that we propose in this paper. Later, we present the reasons for such proposed normalization, and compare our classification results to the state-of-the-art.

A processing is performed on the raw ECG signal (image provided by the HS) from which data are generated for training and classification by ML algorithms, and includes the fiducial point detection algorithm developed for this specific purpose.

Figure 4 illustrates the steps of the method, which are detailed below.

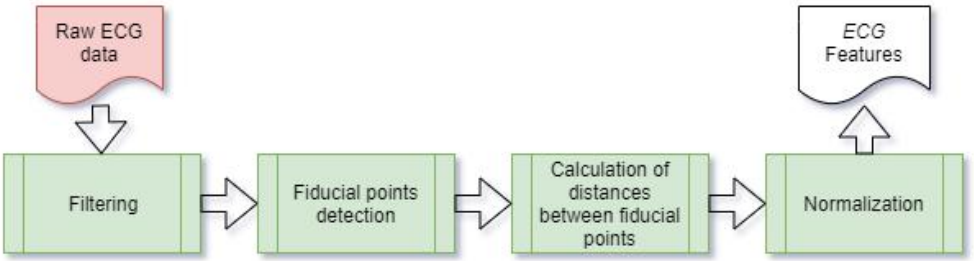

Figure 4- ECG features extraction

Filtering

Based on the raw ECG data, this step aims to reduce high-frequency components that hinder the detection of fiducial points, such as noise associated with the electrical network and transmission of electrical signals. According to [22], the normal frequency range of an ECG is between 1 and 40 Hz.

$E_0[t]$ represents the raw ECG data and stores values of electrical potentials in mV (representing the amplitudes of an ECG)

temporally indexed (the ECG sample rate enables a relationship between a vector index and a relative value of time). This vector contains a number of cycles determined by its duration and the individual's average heart rate. As an example, if $E_0[t]$ represents two minutes of an ECG at a 60 bpm average heart rate, $E_0[t]$ will contain 120 cycles.

Having received an $E_0[t]$ vector , the signal is processed, so that possible noises that distort the exact positions of the fiducial points are attenuated. A low-pass filter, such as Butterworth filter, is required. Figure 5 shows the effect of the application of the indicated filter, which attenuates the various peaks and valleys that represent the signal noise. The filter also enables a simpler and accurate detection of such fiducial points. After the filtering process, $E_0[t]$ becomes vector $E_1[t]$, which represents the filtered ECG signal storing its amplitude values.

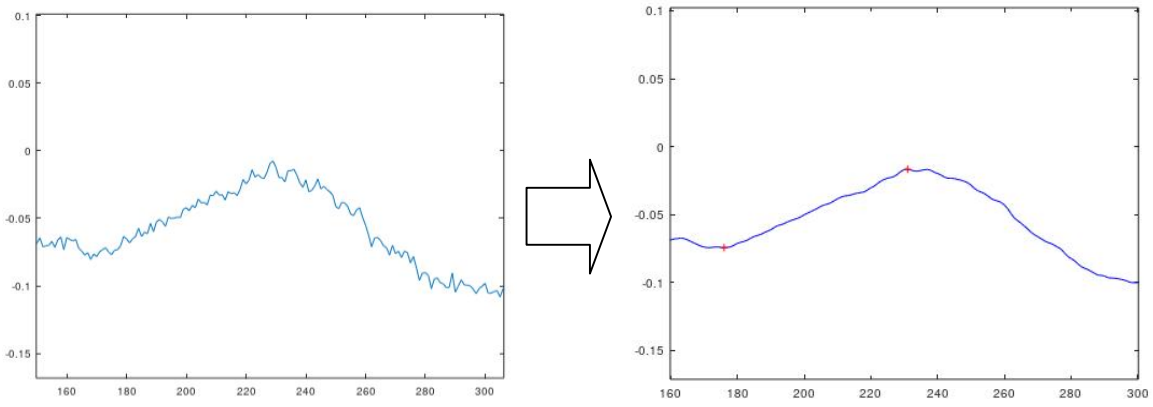

Figure 5- Effect of application of the "Butterworth" filter

Fiducial points Detection

The fiducial point detection algorithm uses MatLab's findpeaks function as an initial step, which indicates a series of peaks and valleys in the analyzed signal. IDX vector is obtained; it contains the values of the $E_1[t]$ indices that represent the moments at which the peaks and valleys were found in $E_1[t]$. Figure 6 shows a filtered cycle (in blue) and the peaks and valleys found (in red).

Some fiducial points are represented by more than one peak / valley detected (e.g., points *LP*, *P* and *TP* in the figure). In such a case, the fiducial point detection algorithm assumes the average value of the peaks / valleys that are in the vicinity of the fiducial point searched. If the point is at the ends of the analyzed cycle (points *LP* and *TP*), the peak / valley considered will be the most internal one (in the case of *LP*, the one on the right, and, in the case of *TP*, the one on the left).

The algorithm starts its analysis at the point normally associated with the largest amplitude of a cardiac cycle, i.e., at peak *R*. From that fiducial point, a search for the 3 points on the left and, subsequently, the 3 points on the right is conducted. The search is based on the relative position between neighboring points (*Q* is below *R*, *P* is above *Q*, etc.).

Finally, the algorithm selects appropriate elements *IDX* and creates *LP[c]*, *P[c]*, *Q[c]*, *R[c]*, *S[c]*, *T[c]* and *TP[c]* vectors, where *LP*, *P*, ..., *TP* are index vectors of $E_1[t]$ for each fiducial point detected, and *c* is the cardiac cycle index considered.

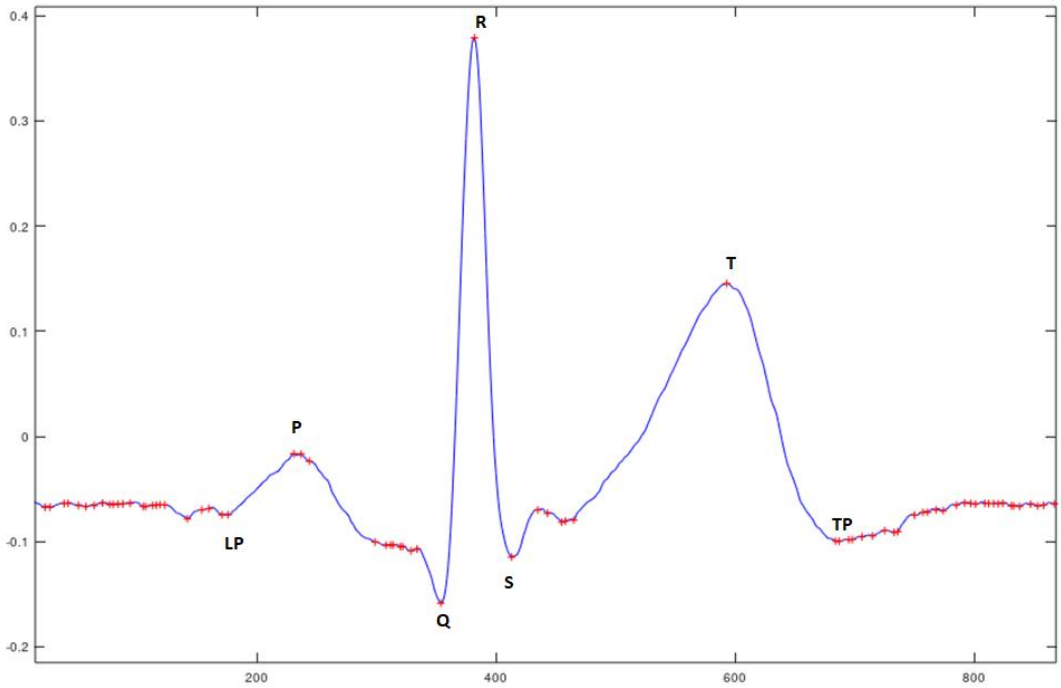

Figure 6- Filtered ECG signal with detected peaks and valleys

Some fiducial points are represented by more than one peak / valley detected (e.g., points *LP*, *P* and *TP* in the figure). In such a case, the fiducial point detection algorithm assumes the average value of the peaks / valleys that are in the vicinity of the fiducial point searched. If the point is at the ends of the analyzed cycle (points *LP* and *TP*), the peak / valley considered will be the most internal one (in the case of *LP*, the one on the right, and, in the case of *TP*, the one on the left).

The algorithm starts its analysis at the point normally associated with the largest amplitude of a cardiac cycle, i.e., at peak *R*. From that fiducial point, a search for the 3 points on the left and, subsequently, the 3 points on the right is conducted. The search is based on the relative position between neighboring points (*Q* is below *R*, *P* is above *Q*, etc.).

Finally, the algorithm selects appropriate elements *IDX* and creates *LP[c], P[c], Q[c], R[c], S[c], T[c]* and *TP[c]* vectors, where *LP, P, ..., TP* are index vectors of $E_1[t]$ for each fiducial point detected, and *c* is the cardiac cycle index considered.

Calculation of distances between fiducial points

Using index vectors *LP[c], P[c], Q[c], R[c], S[c], T[c] e TP[c],* and vector $E_1[t]$, which represents the filtered ECG signal, we extract the values of the distances shown in Figure 7, given in terms of "time" by

$$RFP_i(c) = |R(c) - FP_i(c)|, \quad (1)$$

where $1 \leq i \leq 6 \; and \; FPi = [LP, P, Q, S, T, TP]$, and in terms of "amplitude" by

$$RSA(c) = E_1(R(c)) - E_1(S(c)), \quad (2)$$

$$RQA\,(c) = E_1(R(c)) - E_1(Q(c)). \quad (3)$$

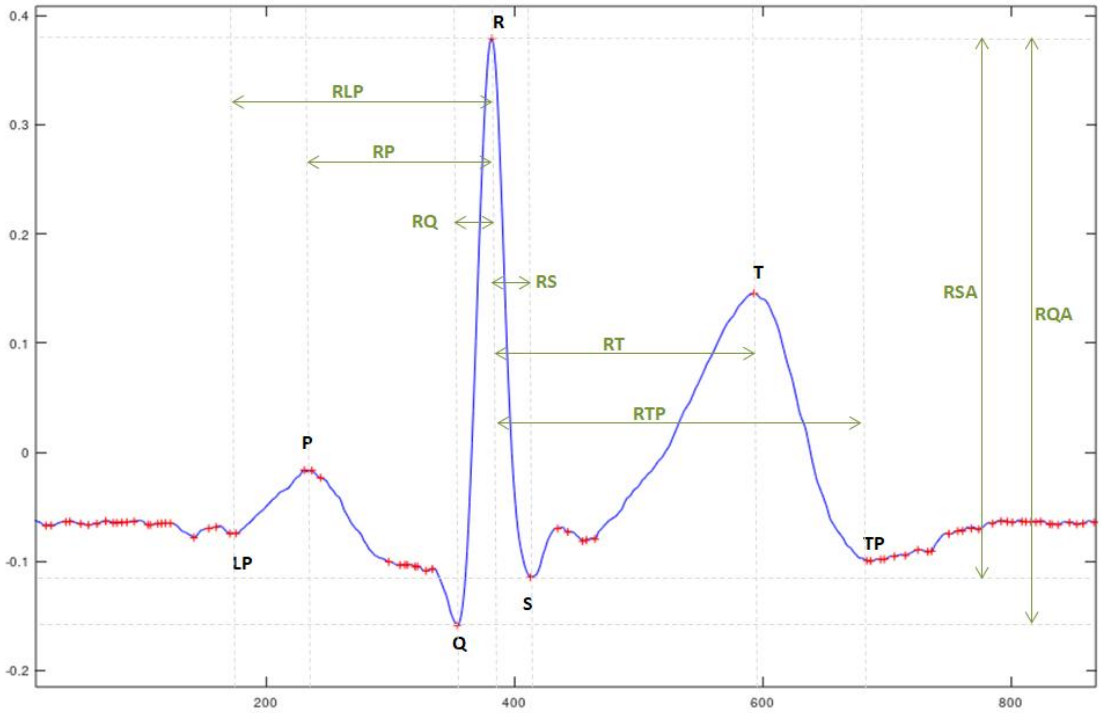

Figure 7- Distances considered for the calculation of ECG features

Normalization

Figure 8 shows a 5 seconds sample typically obtained from an ECG signal that contains several cardiac cycles that will provide, one by one, fiducial point values to be counted for the final calculation of ECG features.

If the cycles are separated and aligned from the same point (usually the R point, which is the most prominent fiducial point on a normal ECG), the total lengths vary from cycle to cycle, due to changes in the heart rate which, as it increases or decreases, compress or expand such lengths, respectively.

Those differences in the length of the cycles (hence, in the lengths of the segments internal to the cycles) can generate distortions in the training of the classifiers, as they indicate different measures for the same characteristic of the cardiac cycle.

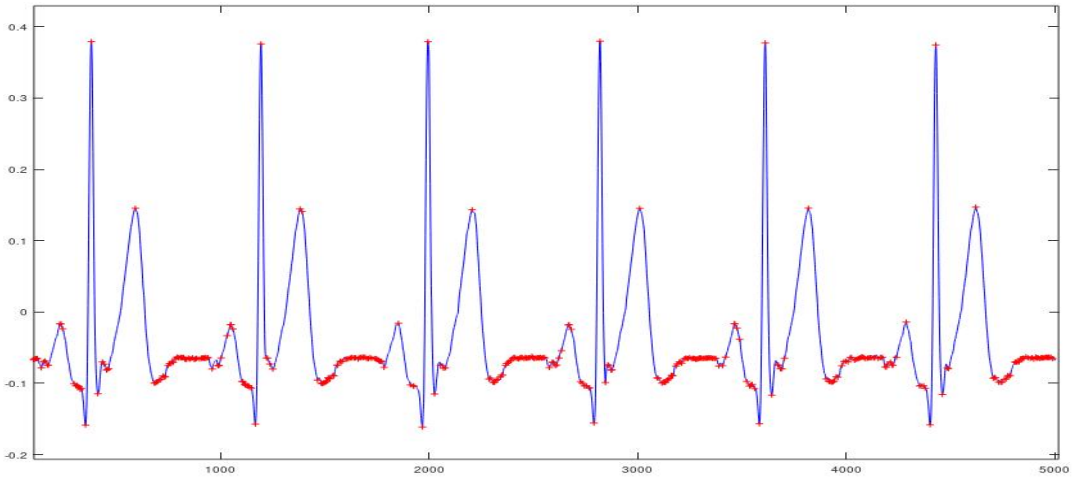

Figure 8- 5 s ECG signal sample

Figure 9 shows a graph of the differences among the various total cycle lengths of the sample in Figure 8. A horizontal normalization of the different cycles analyzed is required for mitigating the effects of heart rate variations that generate different cycle lengths. It is performed by the calculation of the ratio of the previously found distances (in terms of indices) to the total cycle length (distance from the *LP* point to the *TP* point, in terms of indices of each cycle, or simply, *LP-TP* distance).

[13], [14] and [20] used the distances between consecutive *R* points (*RR*) for normalization. Such distances are supported by references highlighted by the authors, however, they may be questioned (see [23]). Although variations in the temporal distances between the fiducial points (e.g., QRS) occur in the same direction of the variation in the RR distance (when it increases, the others also increase), they are not proportional to the variation of that distance.

The adoption of the *LP-TP* distance for normalization aimed at neglecting the effects of heart rate variation perceived in the portion of the ECG between the *TP* point of a cycle and the *LP* point of the next cycle (i.e., a section of the ECG that is outside the cardiac cycle standardized here). Although no references were found to support the use of this measure, the experimental results reported in the next section indicate this choice was favorable.

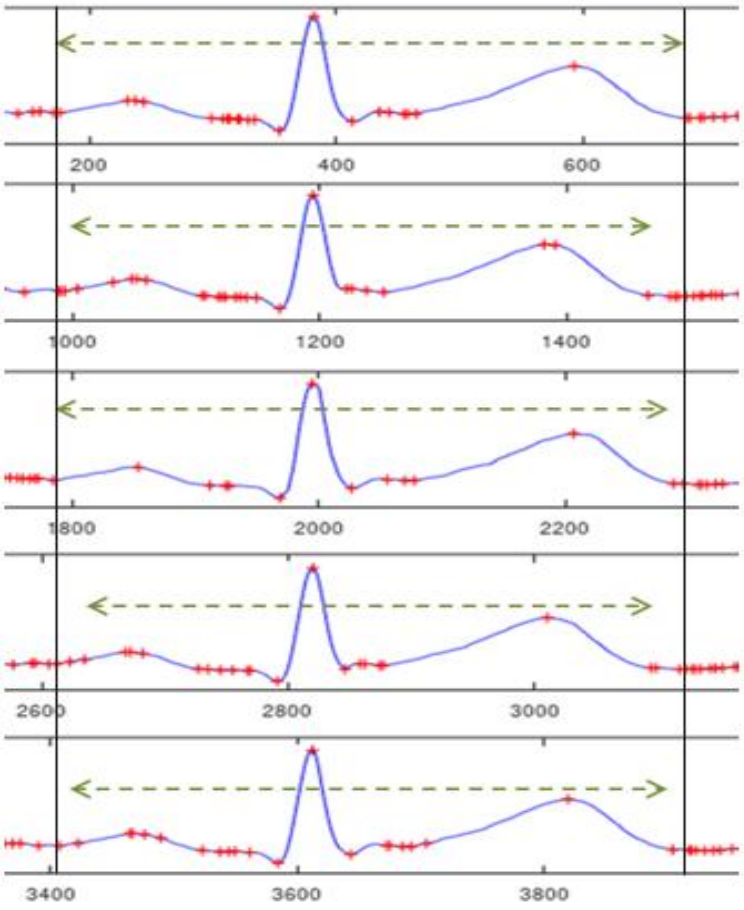

Figure 9- Differences in cycle lengths

Considering equation (1), the normalization calculations finish the process of extraction of ECG features as

$$f_i(c) = \frac{RFPi(c)}{D(c)}, \quad (4)$$

where $D(c) = TP(c) - LP(c)$ and the equations represent the ratios between the segments and the longest distance in the cycle, as explained above.

The two other features are given in mV (ECG amplitude) and do not undergo horizontal normalization. The equations that represent them are

$$f_7(c) = RSA(c) \quad (5)$$

and

$$f_8(c) = RQA\ (c). \quad (6)$$

Finally, vectors $f_1[c], f_2[c], \ldots, f_8[c]$ represent the features extracted for all the cycles considered. Although such features are derived from the same basic ideas of [13], their ways of obtaining differ in the detection of fiducial points and normalization.

3.4.2. Training classifier method

The following two machine learning classifiers were used in our study: Adaptive Boosting and NNS (Nearest Neighbor Search).

AdaBoost

Adaptative Boosting, or simply AdaBoost, is an ensemble that represents one of the main results of the application of the boosting technique. Its main objective is to minimize, for a binary classification, the exponential loss function [17] given by

$$lo_{exp}(Class|D) = E_{x\sim D}\left[e^{-f(x)Class(x)}\right], \quad (7)$$

where lo_{exp} is the exponential loss function, *Class* is the classifier to be trained considering the sample distribution *D*, *x* is an example subjected to the training process, and *f* is the function that returns the label of *x*.

It has been used for the classification of physiological signals and shown good results [24]. Details on this technique can be found in [17].

Nearest Neighbor Search – NNS

The search for the nearest neighbor (NNS) or, in more generic contexts, the neighbor of proximity order k (kNN), is based on the idea that objects that are similar in the entry space of a classification process are also similar in the exit space [17].

NNS is an example of a classifier that uses non-parametric learning rules [16], i.e., rules that are not based on a parametric model of the relationship between inputs and outputs. Instead, it considers training examples that have close values tend to have similar labels.

From a geometric point of view, NNS aims at finding, in an n-dimensional space, a point, among several that belong to the same set of points, that is closer to a certain reference position. It was used by [13], for example, who also focused on the use of electrocardiography for authentication..

A specific process of training is employed for each of the two classifiers chosen, as demonstrated below.

AdaBoost Training

AdaBoost is trained using the MatLab *fitcensemble* function. AdaBoostM1 values are considered as parameters for the choice of the classifier type, whereas the feature values matrix, detailed in section 4, are considered as examples for training. After training, the function returns data structures that contain all classifier parameters that must be stored in the ADB.

NNS Training

NNS training consists exclusively in determining the values of the averages of the features for all cycles considered in the analysis, i.e.,

$$\mu_{fj} = \frac{1}{NC}\sum_{c=1}^{NC} f_j\ (c), \tag{7}$$

where $1 \leq j \leq 8\ e\ \mu_{fj}$ is the average of the "j-th" feature considering all training cycles (*NC*).

After the calculations, the mean values are stored in the ADB as classifier parameters.

3.4.3. Authentication Method

Authentication with AdaBoost

Authentication with this classifier is based on the application of MatLab *predict* function, which uses the classifier parameters obtained from ADB and a vector of features for authentication. After classification, the indicated functions return a positive or negative result for signaling authentication.

Authentication with NNS

In this process, all time-based fj features (horizontal ECG distances) are compared to the μ_{fj} averages stored in the ADB, based on specific tolerances Tol_j. The total number of successes in the comparisons will be recorded in variable *Hit*. For the user to be considered authenticated, the number of correct answers must be greater than or equal to a minimum value of minutes and the two amplitude-based features (f_7 and f_8) must be within the appropriate range of values (considering the stored averages and tolerances).

4. EVALUATION OF THE AUTHENTICATION PROCESS BASED ON ELECTROCARDIOGRAPHY

4.1. INITIAL CONSIDERATIONS

Some activities, described in the following items, were undertaken for validating the authentication process based on electrocardiography. A personal computer with an Intel i5 processor, 16 GB of main memory and 1 TB of disk was used.

Cardiac cycles (*c*) are considered curves that represent parts of an ECG between a fiducial point *LP* and the subsequent fiducial point *TP*, i.e., waveforms containing points *LP*, *P*, *Q*, *R*, *S*, *T* and *TP*. After recognizing the curves, the ECG feature extraction process detects the fiducial points of each cycle and extracts the 8 features ($f1, f2, \ldots, f8$) used in the authentication process (each cycle is represented by its features).

The repository with 108 users developed by [25] was used (see Table 1). As shown in the table, the useful total of users considered is lower than the gross total, since some capture processes failed to read the signals from the ECG sensors, and hampered the use of signals of some individuals [25].

Table 1 - Characteristics of the ECG base used.

Feature	Value
Gross total of users	132
Nu_{Tot}	108
Number of ECGs per user	1
Useful total of men	56
Useful total of women	52
Age range 1 - (18 to 35 years)	98
Age range 2 - (36 to 66 years)	10
$\Delta t_{ECG\ Min}$	300s
Sampling rate	500 Hz
Heart rate (lower than 60)	6
Heart rate (60 to 120)	102
Heart rate (over 120)	0

During the initial analyses, the quality of the ECG of the base chosen was quite variable (probably due to cardiac problems of the participants or problems in the capture of electrical signals). It hampered the detection of fiducial points by the method of extraction of features, thus causing several of such points to not be detected and forcing the cycles to which they belong to be considered unrecognized and discarded from the analysis.

4.2. PREPARATION OF THE FEATURES USED IN THE AUTHENTICATION MODEL

Once the detection of the fiducial points and consequent recognition of the users' cardiac cycles are completed, the features of all cycles recognized by the user are extracted. Such cycles are represented by

$$c_{i,u} = [f_{1,i,u}, f_{2,i,u}, \dots f_{8,i,u}], \tag{8}$$

where $c_{i,u}$ represents a user's "i-th" cardiac cycle and $f_{j,i,u}$ is the "j-th" feature of $c_{i,u}$ cycle.

The joining of all $c_{i,u}$ of a user yields the matrices (M_u) of features of all cycles recognized by the user. Such matrices have varying numbers of rows per 8 columns (NC_u by Nf_{Tot}), and can be represented by

$$M_u = \begin{bmatrix} f_{1,1,u} & f_{2,1,u} & \cdots & f_{8,1,u} \\ f_{1,2,u} & f_{2,2,u} & \cdots & f_{8,2,u} \\ & \ddots & & \\ f_{1,NC_u,u} & f_{2,NC_u,u} & \cdots & f_{8,NC_u,u} \end{bmatrix}_{NC_u \times Nf_{Tot}} \tag{9}$$

where NC_u is the number of cardiac cycles recognized per user and Nf_{Tot} is the total number of features per cycle used in this study.

The M_u matrices are concatenated for the application of the random sampling process, addressed in Section 4.3, thus creating matrix *M*, given by

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{Nu_{Tot}} \end{bmatrix}_{NC_{Tot} \times Nf_{Tot}} . \tag{10}$$

A column matrix (*V*) is also created and represents the identification numbers of the users who supplied the ECGs that originated the features used, so that the features of each *M* line belong to a user's ECG cycle indicated in the corresponding line of *V*. The column matrix is represented by

$$V = \begin{bmatrix} 1 \\ \vdots \\ 108 \end{bmatrix}_{NC_{Tot} \times 1} . \tag{11}$$

4.3. RANDOM SUB-SAMPLING PROCESS FOR EXPERIMENTATION

After the creation of the matrices, a random subsampling technique consisting of 100 validation sessions per user and derived from the holdout technique indicated in [26] is adopted. It aims to validate the authentication process based on electrocardiography and using the features obtained in the extraction process. One of its main objectives is to analyze which classifier (AdaBoost or NNS) has better conditions to be used in the authentication process.

The analysis of the classification process considers four metrics, namely accuracy (*Acc*), precision (*Pre*), sensitivity (*Sen*) and F1-score (*F1s*), which are based on the quantities of true positives (*tp*), true negatives (*tn*), false positive (*fp*) and false negative (*fn*), and calculated by the following equations:

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} , \tag{12}$$

$$Pre = \frac{tp}{tp + fp} , \tag{13}$$

$$Sen = \frac{tp}{tp + fn}, \quad (14)$$

$$F1s = 2 \cdot \frac{Pre \cdot Sen}{Pre + Sen}. \quad (15)$$

The best classifier found according to those metrics is used in the authentication protocol.

Note that we first trained each classifier using a random set of input signals from the dataset, and later we evaluated its performance using a different set. We emphasized the idea of not adjusting classifier parameters based on features that were later used for the performance evaluation, in order to avoid over fitting. Furthermore, we repeated this process of training and testing the classifiers several times, until the variation of evaluated metrics were below a specified tolerance, according to the random resampling method described in [26]. These procedures lead to our performance metrics estimates while avoiding over and under fitting, as the final metrics are stable after a large number of sessions, i.e., new training/testing sessions do not increase or decrease the estimates.

The validation sessions of each user employ a random subsample ($SAM_{u,s}$) of the features of matrix *M* defined above, where *u,s* define a session *s* of a user *u*. Each $SAM_{u,s}$ is composed of the following matrices:

- *M1_training u,s* – a matrix randomly obtained from *M* that contains 70% of the cycles of *u* and serves for the training of classifiers;
- *M1_ validation u, s* – a matrix obtained from *M* that contains the remaining 30% of the cycles of *u* and is used for the validation or classification performed by the classifiers;
- *M0_training u,s* – a matrix randomly obtained from *M* that contains 70% of user cycles other than *u* and is used to train classifiers;
- *M0_ validation u, s* – a matrix randomly obtained from *M*, of same dimensions of *M1_ validation u, s*, formed by user cycles other than *u*, and used for the validation or classification made by the classifiers.

The "70% x 30%" proportions applied to this process are roundings of the "2/3 x 1/3" proportions commonly found in the literature on classifier training and testing processes (as in [26] and [14]). Algorithm 1 shows the steps taken in the random subsampling process and, in summary, performs the following tasks:

1. Data separation:
   - obtaining of matrices $M1\_training_{u,s}$, $M1\_validation_{u,s}$, $M0\_training_{u,s}$ and $M0\_validation_{u,s}$;
2. Classifier training:
   - training is performed for each of the 2 classifiers, with the matrices previously obtained (*M1_training* and *M0_training*);
3. Validation or classification:
   - similarly to training, validations are performed with each trained classifier and features of M1_validation and M0_validation;
4. Collection of results:
   - The total number of hits and errors for all samples is counted for each validation towards verifying whether the process "hits" or "misses";
   - The errors and successes are defined as follows:

a. true positive - *tp* - classifier correctly indicates the features of the standard used in its training;

b. true negative - *tn* - classifier correctly indicates features that do not belong to the standard used in its training;

c. false positive - *fp* - classifier erroneously indicates the features of the standard used in its training;

d. false negative - *fn*- classifier erroneously indicates features that do not belong to the standard used in its training;

5. Obtaining of analysis metrics:
   - According to the errors and hits collected, the analysis metrics previously defined are counted: Accuracy, Precision, Sensitivity and F1-score.

Algorithm 1 - Steps of the random subsampling process

START
Total number of sessions (Ns_{Tot}) = 100
Total number of users (Nu_{Tot}) = 108
Classifiers set (Set_{Class})= [Ada, NNS]
Metrics set (Set_{Metr})= [Acu, Pre, Sen, F1s]
For each user (u) from 1 to Nu_{Tot}, do:
For each session (s) from 1 to Ns_{Tot}, do:
Retrieve M1_training$_{u,s}$, M1_ validation$_{u,s}$,M0_training $_{u,s}$ and M0_ validation $_{u,s}$
For each classifier (Class) ∈ Set_{Class} , do:
Train Class using M1_training$_{u,s}$ e M0_training $_{u,s}$
For each cicle (c) ∈ [M1_ validation$_{u,s}$,M0_ validation $_{u,s}$]
Classify c as [true_positive (tp), false_positive (fp), true_negative (tn) or false_negative (fn)]
For each metric (Metr) ∈ Set_{Metr}, do:
Calculate Metr using amount of tp, fp, tn and fn
END

4.4. VERIFICATION OF THE "STABILITY" OF THE METRICS VALUES OF ANALYSIS

After all sessions had been run by the steps listed above (100 random subsampling sessions per classifier per user), they were organized, for each user, into 10 groups ($Gs_{k,u}$) of validation sessions of the random subsampling technique. The quantity of $Gs_{k,u}$ ranged from 10 to 100 sessions, with steps of 10, i.e., $Gs_{k,u} \in [10, 20, \ldots, 100]$.

The division into groups of varying sizes enabled comparisons of the average values obtained from groups whose sizes gradually increase for checking the “stability" of the average value of the metrics (from a certain group size, the average value does not change substantially and can be adopted as the average value of the metric in the considered classifier).

After the groupings ($Gs_{k,u}$) , average value μ and standard deviation σ of each analysis metric (*Acu*, *Pre*, *Sen* and *F1s*) of each classifier (*Ada* and *NNS*) were calculated considering each $Gs_{k,u}$.

The estimates of μ and σ were obtained by

$$\mu_{Class,Metr,k,u} = \frac{1}{k}\sum_{s=1}^{k} Metr_{Class,u,s} \tag{16}$$

and

$$\sigma\left(\mu_{Class,Metr,k,u}\right) = \sqrt{\frac{1}{k-1}\sum_{s=1}^{k}\left(Metr_{Class,u,s} - \mu_{Class,Metr,k,u}\right)^2}\,. \tag{17}$$

After all mean values and standard deviations of the 8,640 possible combinations (2 classifiers, 4 metrics, 10 groups of sessions and 108 users) were calculated, the means of μ and σ are calculated considering all users. After the stability of the average value of the metrics had been checked from the calculations, the graphs of Figures 10 and 11 were obtained.

The graphs show the stability of the mean values of each metric, in each classifier, according to the parallelism of their curves (almost straight) in relation to the abscissa axis. It signalizes such values do not tend to change substantially in function of the increasing k. Ideally, the higher the value of this variable, the more representative the values of the averages of the analyzed metrics. However, the small variations from the increase in k are estimated to justify the interruption of its evolution in 100.

In view of the previously perceived stability and the fact that all metrics for all classifiers are calculated in up to 100 sessions (Ns_{Tot}), it is valid to assume the averages of the metrics calculated with this maximum number of sessions are, indeed, stabilized representations and can be adopted as standard mean values for the metrics in the various classifiers.

4.5. ANALYSIS OF THE GENERAL VALUES OF THE METRICS

According to the previous conclusion on stability, only the averages of the metrics calculated from the total sessions performed ($k = Ns_{Tot}$) were adopted for the other analyses conducted. For simplicity, such averages calculated with $k = Ns_{Tot}$ are hereinafter referred to only as metrics (*Acu, Pre, Sen, F1s*) and, likewise, the averages of standard deviations are called deviations standard (σ).

Table 2 shows a general comparison of the metrics and the respective standard deviations.

Table 2 - values of the analysis metrics and the standard deviations in all classifiers used.

Class	Acu ± σ	Pre ± σ	Sen ± σ	F1s ± σ
Ada	0,807 ± 0,04	0,842 ± 0,057	0,615 ± 0,079	0,69 ± 0,074
NNS	0,753 ± 0,052	0,847 ± 0,085	0,51 ± 0,104	0,6 ± 0,1

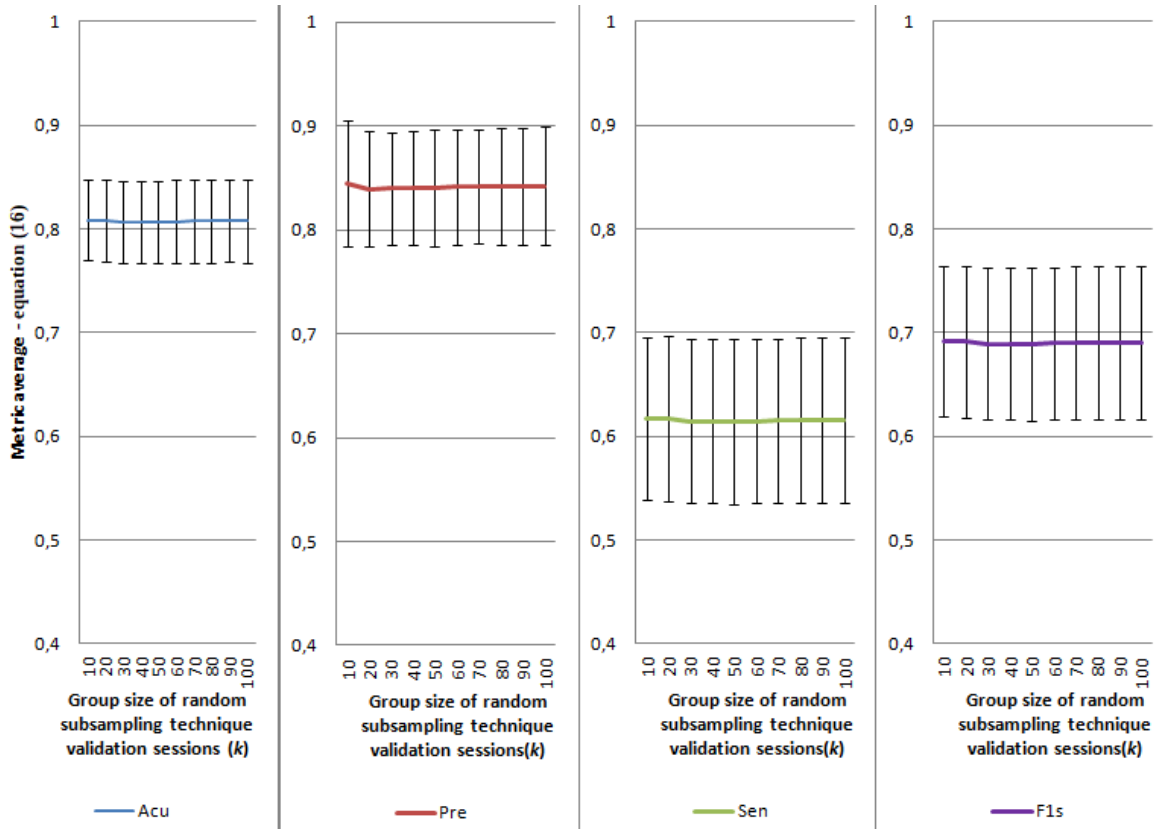

Figure 10 - AdaBoost - graphs of the stability behavior of the mean values of the analysis metrics

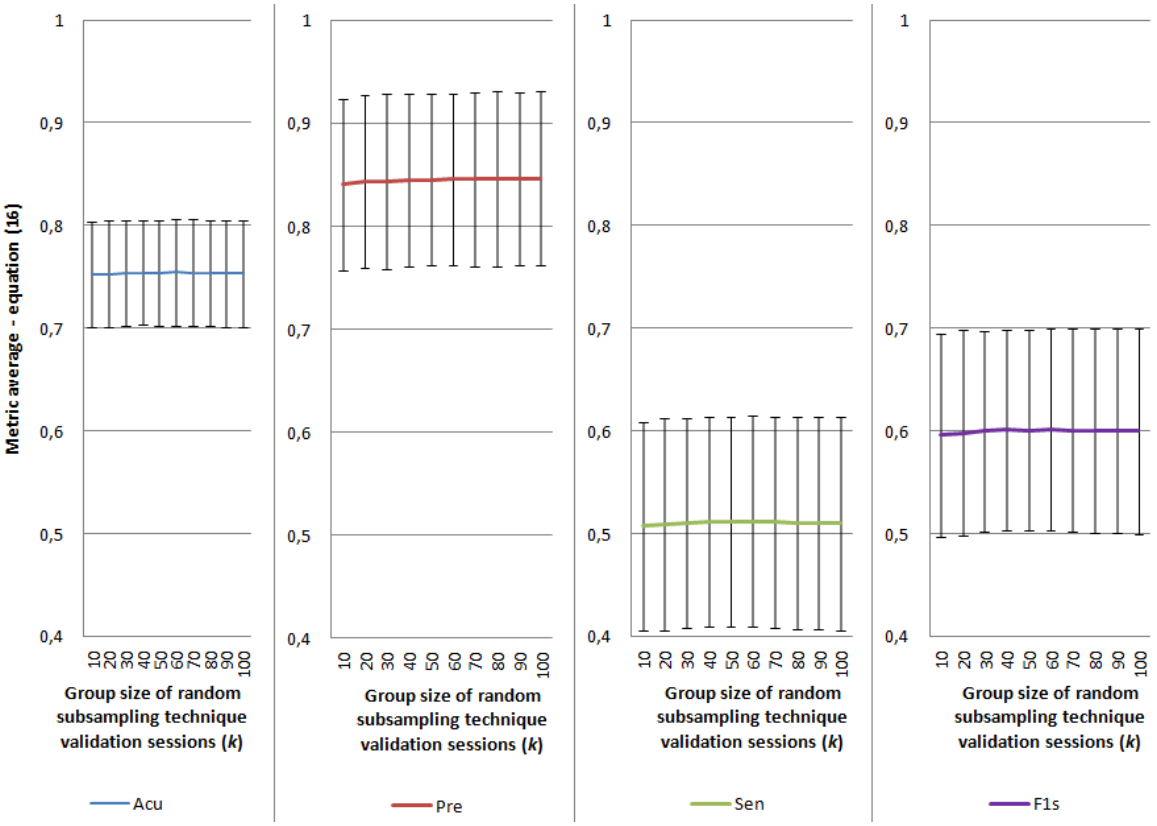

Figure 11- NNS - graphs of the stability behavior of the mean values of the analysis metrics

AdaBoost obtained the best values for *Acu*, *Sen* and *F1s*, and for their respective standard deviations. Regarding Pre and its respective standard deviation, this classifier ranked second.

Towards the isolation of the problems that generated recognition of cycles lower than the maximum possible ($NC_u < NC_{Max}$), the values in Table 3 are similar to those of Table 2 for cases in which $NC_u = NC_{Max}$.

Table 3 - values of the analysis metrics similar to that of Table 2 for the cases in which $NC_u = NC_{Max}$.

Class	Acu ± σ	Pre ± σ	Sen ± σ	F1s ± σ
Ada	0,920 ± 0,032	0,962 ± 0,038	0,840 ± 0,064	0,893 ± 0,052
NNS	0,843 ± 0,033	0,995 ± 0,008	0,689 ± 0,065	0,791 ± 0,051

Although the values in Table 2 are lower than those in Table 3, the main result of the comparison between the classifiers is very close to that from the previous comparison, i.e., AdaBoost obtained the best values in most metrics and standard deviations analyzed.

Despite the previous conclusions representing some of the main results from the analyses conducted, some considerations on the classifiers applied to the several users of the dataset chosen must be made. The analyses that support such considerations were based on the data of Tables 4 and 5, which contain sample data from the 108 users of the ECG database. The columns in the tables follow a color scale based on the idea of a heat map. The colors vary, in an ascending order, from the lowest to the highest value according to the legend:

Lowest values — Highest values

Table 4- AdaBoost - Heat map representing the values of the analysis metrics and standard deviations for 17 users ordered by the number of recognized cycles.

NC_u	u	Acu ± σ	Pre ± σ	Sen ± σ	F1s ± σ
3	104	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
11	52	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
13	76	0,519 ± 0,051	0,150 ± 0,359	0,040 ± 0,099	0,063 ± 0,152
18	45	0,500 ± 0,000	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,000
36	35	0,690 ± 0,062	0,990 ± 0,100	0,379 ± 0,123	0,538 ± 0,138
43	61	0,678 ± 0,067	1,000 ± 0,000	0,357 ± 0,134	0,512 ± 0,149
82	36	0,895 ± 0,041	1,000 ± 0,000	0,790 ± 0,082	0,880 ± 0,053
103	27	0,937 ± 0,056	0,999 ± 0,007	0,874 ± 0,111	0,928 ± 0,069
127	4	0,635 ± 0,036	1,000 ± 0,000	0,271 ± 0,072	0,421 ± 0,091
143	25	0,907 ± 0,025	1,000 ± 0,003	0,813 ± 0,049	0,896 ± 0,030
179	15	0,800 ± 0,030	0,997 ± 0,008	0,602 ± 0,061	0,749 ± 0,047
196	70	0,872 ± 0,024	0,999 ± 0,004	0,744 ± 0,047	0,852 ± 0,031
248	84	0,953 ± 0,016	1,000 ± 0,001	0,907 ± 0,032	0,951 ± 0,018
292	92	0,978 ± 0,014	1,000 ± 0,002	0,955 ± 0,027	0,977 ± 0,014
300	26	0,931 ± 0,021	0,999 ± 0,004	0,863 ± 0,041	0,926 ± 0,024
300	41	0,987 ± 0,010	1,000 ± 0,001	0,975 ± 0,020	0,987 ± 0,011
300	51	0,974 ± 0,011	1,000 ± 0,002	0,948 ± 0,023	0,973 ± 0,012

Table 5- NNS - Heat map representing the values of the analysis metrics and standard deviations for 17 users ordered by the number of recognized cycles.

NC_u	u	Acu ± σ	Pre ± σ	Sen ± σ	F1s ± σ
3	104	0,645 ± 0,228	0,290 ± 0,456	0,290 ± 0,456	0,290 ± 0,456
11	52	0,557 ± 0,119	0,225 ± 0,417	0,117 ± 0,239	0,148 ± 0,286
16	64	0,664 ± 0,237	0,330 ± 0,473	0,330 ± 0,473	0,330 ± 0,473
36	35	0,561 ± 0,087	0,458 ± 0,493	0,127 ± 0,174	0,188 ± 0,240
43	61	0,803 ± 0,103	0,989 ± 0,041	0,612 ± 0,201	0,736 ± 0,173
82	36	0,785 ± 0,078	1,000 ± 0,000	0,570 ± 0,155	0,714 ± 0,132
102	83	0,577 ± 0,048	0,970 ± 0,171	0,155 ± 0,095	0,256 ± 0,138
117	87	0,544 ± 0,034	0,892 ± 0,266	0,092 ± 0,066	0,161 ± 0,107
136	6	0,656 ± 0,054	0,989 ± 0,035	0,316 ± 0,108	0,469 ± 0,124
196	70	0,633 ± 0,034	0,998 ± 0,011	0,267 ± 0,069	0,416 ± 0,086
212	90	0,721 ± 0,047	0,980 ± 0,028	0,452 ± 0,094	0,613 ± 0,092
270	9	0,605 ± 0,027	1,000 ± 0,000	0,211 ± 0,055	0,345 ± 0,076
292	92	0,911 ± 0,034	0,991 ± 0,012	0,830 ± 0,067	0,902 ± 0,041
300	7	0,938 ± 0,025	0,996 ± 0,007	0,880 ± 0,049	0,934 ± 0,029
300	41	0,941 ± 0,025	0,998 ± 0,005	0,883 ± 0,049	0,936 ± 0,029
300	51	0,953 ± 0,021	0,993 ± 0,009	0,912 ± 0,043	0,950 ± 0,024
300	71	0,973 ± 0,018	0,997 ± 0,006	0,948 ± 0,036	0,972 ± 0,019

Below are some considerations on the values of the users' metrics per classifier:

Regarding AdaBoost:

- Users of a first range of recognized cycle values ($NC_u < 11$) show very poor values (Acu = 0.5 and other Metr = 0);
- in a second range, with cycle values recognized according to $11 \leq NC_u \leq 18$, although the values of all metrics slightly increased, they are still considered bad (several $Metr < 0{,}6$);
- for $18 < NC_u \leq 127$, Acu, Sen and F1s considerably improved (metrics ranged between 0.6 and 0.8) and Pre largely increased (several values reached 1);
- for $127 < NC_u \leq 300$, practically all metrics showed values above 0.8 (with some exceptions that regressed to values below 0.5) and, in particular, Pre continued showing high results (several values equal to or close to 1).

Regarding NNS:

- Almost all users of a first range of recognized cycle values ($NC_u < 16$) showed very bad values (Acu = 0.5 and other Metr = 0);
- in the second range ($16 \leq NC_u \leq 102$), although the values of all metrics improved (several values greater than 0.7), a high occurrence of very bad values (lower than 0.5) was observed in Sen and F1s;
- for $102 < NC_u \leq 270$, Acu, Sen and F1s marginally improved (almost all Acu values ranged between 0.55 and 0.8, and many Sen and F1s values were below 0.5) and substantially improved (several values were close to or equal to 1);
- for $270 < NC_u \leq 300$, Acu, Sen and F1s improved and showed several values higher than 0.8, and Pre maintained its excellent results.

The values in Tables 3 and 4 and the different perceptions mentioned have proven

AdaBoost yields better results than NNS for this classification problem.

4.6. ANALYSIS OF THE NUMBER OF CYCLES FOR A TRAINING

Regarding the number of cycles necessary for a classifier to be considered well trained, the two classifiers evaluated obtained better metrics values as the number of cycles used in their training increased. This result is perceptible in the tables, with a few exceptions.

Such a conclusion corroborates the general idea in the machine learning area that points to a proportionality between the quality of a training and the sample size used in it.

The limits imposed by the chosen database in the present study hampered the analysis of whether values greater than 300 cycles can provide better results for metrics than those indicated for $NC_u = 300$. Anyway, the results in Table 4 can be considered good, which leads to the conclusion at least 300 cycles should be used for the training of the chosen classifier.

4.7. COMPARISON WITH RESULTS OF THE LITERATURE

In this subsection, we present a comparative table involving the characteristics of the proposals, as well as a synthesis of the results. The main differences pointed out in the table refer to the normalization processes of the features, the ECG databases used and the final values of the metrics. Table 4 shows the comparison among the four articles and our proposal.

Table 4- Comparison among the studies analyzed

Features / studies	[14]	[13]	[15]	[20]	Our proposal
Authentication factors	ECG	ECG	ECG	ECG and fingerprint	ECG
Approach	Detection of fiducial points	Detection of fiducial points	Walsh-Hadamard transform	Detection of fiducial points	Detection of fiducial points
ML Technique	SVM	NNS	kNN (k=1)	SVM (for ECG)	AdaBoost and NNS
Number of ECG features	31	8	Untreated	8	8
Are ECG data normalized?	Untreated	Normalization based on R-R	Not applicable	Normalization based on R-R	Normalization based on LP-TP
ECG database used	[27], [28]	[27], [28], [29], [30]	[28]	[27], [28], [29], [30]	Base obtained from research at UnB [25]
Overall number of users analyzed	66	73	10	73	108
Max. Accuracy	99.38%	81,82%	94,79%	100%	92,0%
Max. Precision	-	-	-	-	99,5%
Max Sensitivity	-	-	-	-	84,0%
Max F1-Score	-	-	-	-	89,5%

As we can see, our proposal used a database composed of a larger amount of ECG than the other works. In addition, in our work, 4 metrics were analyzed compared to a single metric used in the other proposals and an innovative normalization technique was applied using the *LP-TP* distance.

5. CONCLUSIONS

User authentication in mobile cloud computing environments represents a challenge. Several protocols have been developed towards overcoming it, and those based on biometric authentication have excelled, since they enable the use of data intrinsic to the human being who intends to use the services of a given provider.

This article has addressed a proposal and evaluation of a protocol that uses biometric signals based on electrocardiograms. Signal processing and machine learning techniques were applied to a population of 108 users. A protocol that acts on an MCC architecture in an integrated manner and performs operations in a non-intrusive and continuous way, thus improving information security, was obtained.

The article also reported on the implementation of a new algorithm that detects fiducial points based on a pre-defined pattern of the cardiac cycle. Such points are used in the calculation of various features, and the results of the application of the algorithm showed promising.

Another contribution of this research is the proposal and evaluation of a new form of normalization of the temporal characteristics of the cardiac signal, which provides good quality features for ML processes. Although the strategy could not be directly compared with other similar ones - the ECG base used is different from that applied in other studies - its good results proved it can replace others.

The use of AdaBoost for the classification of electrocardiography signals towards benefiting user authentication is also innovative, and, to the best of our knowledge, no other application of this ML technique has been developed for such a purpose of user authentication.

As a last contribution, the application of the random subsampling technique for the experimentation necessary for classification processes and that obtains analysis metrics (accuracy, precision, sensitivity and f1-score) of the authentication process, leading to objective results, hence, relevant conclusions.

Work in progress has included the ML techniques addressed here, towards a multifactorial authentication protocol that employs other classifiers based on MCC architecture. Further research will involve:

- Use of the classifiers applied by changing the features obtained from the ECG;
- Comparison of results from the same ECG and from the application of different measures (*R-R*, *LP-TP*) for the normalization process of the features;
- Increase in the quality of the fiducial point detection algorithm towards a larger number of recognized cycles;
- Design of a new algorithm for the detection of fiducial points that does not focus only on ECGs with cycles following other models/definitions.

REFERENCES

[1] - HRESTAK, D., PICEK, S. (2014). "Homomorphic Encryption in the Cloud", MIPRO 2014 - 37th International Convention on Information and Communication Technology, Electronics and Microelectronics.

[2] - JIANG, Q., MA, J., WEI, F. (2016), "On the security of a privacy-aware authentication scheme for distributed mobile cloud computing services", IEEE Systems Journal, Vol. PP, No. 99.

[3] - DINH, H. T., LEE, C., NIYATO, D., WANG, P., (2013). "A survey of mobile cloud computing: architecture,

applications, and approaches". Wireless Communication Mobile Computing, Vol. 13.

[4] - SUO, H., LIU, Z., WAN, J., ZHOU, K., (2013). "Security and privacy in mobile Cloud Computing,", IWCMC - 9th International Wireless Communications and Mobile Computing Conference.

[5] - RYAN, M. D. (2013). "Cloud Computing security: The scientific challenge, and a survey of solutions". Journal of Systems and Software, Vol. 86, No. 9.

[6] - ALIZADEH, M., ABOLFAZLI, S., ZAMANI, ., BAAAHARUN, S., SAKURAI, K. (2016). "Authentication in Mobile Cloud Computing: A Survey". Journal of Network and Computer Applications, Vol. 61, No. C.

[7] - HE, D., KUMAR , N., SHEN, J. (2018). "Efficient Privacy-Aware Authentication Scheme for Mobile Cloud Computing Services". IEEE Systems Journal, Vol. 12, No. 2.

[8] - MOHSIN, J. (2017). "Two Factor Vs Multi-Factor, an Authentication Battle in Mobile Cloud Computing Environments". International Conference on Future Networks and Distributed Systems Proceedings, No. 39.

[9] - JANNATI, B. (2017). "An Improved Authentication Protocol for Distributed Mobile Cloud Computing Services". International Journal of Critical Infrastructure Protection, Vol. 19.

[10] - AUDITHAN, S., VIJAYASARO, V., VIJAYAKUMAR, P., VIJAYAKUMAR, V. (2017). "An Efficient Authentication Scheme for Mobile Cloud Computing Services". Journal of Information Science and Engineering, Vol. 33, No. 3.

[11] - AL-RUBAIE, M., CHANG, J. M. (2016). "Reconstruction Attacks Against Mobile-Based Continuous Authentication Systems in the Cloud". Information Forensics and Security, IEEE Transactions On, Vol. 11, No. 12, pp. 2648 – 2663.

[12] - QOSENIA, A., SUR-KOLAY, S., RAGHUNATHAN, A., JHA, N. K. (2017) "CABA: Continuous Authentication Based on BioAura". IEEE Transactions on Computers, Vol. 66, No. 5, pp. 759-772.

[13] - ARTEAGA-FALCONI, J. S., OSMAN, H. A., SADDIK, A. E. (2016). "ECG Authentication for Mobile Devices". Instrumentation and Measurement, IEEE Transactions On, Vol. 65, No. 3, pp.591-600.

[14] REZGUI, D., LACHIRI, Z. (2016), "ECG Biometric Recognition Using SVM-Based Approach". Institute of Electrical Engineers of Japan - Transactions on Electrical and Electronic Engineering, Vol. 11, No. S1.

[15] CAMARA, C., PERIS-LOPEZ, P., GONZALEZ-MANZANO, L., TAPIADOR, J. (2017). "Real-Time Electrocardiogram Streams for Continuous Authentication". Applied Soft Computing Journal, Vol. 68, pp. 784-794.

[16] - SIMEONE. O. (2017). "A Brief Introduction to Machine Learning for Engineers". Computer Science, Mathematics, Foundations and Trends in Signal Processing 2017.

[17] - ZHOU, Z. (2012)."Ensemble Methods - Foundations and Algorithms". 1st. edition, Chapman and Hall - CRC.

[18] - FERRAG, M. A., MAGLARAS, L., DERHAB, A. (2019) - "Authentication and Authorization for Mobile IoT Devices Using Biofeatures: Recent Advances and Future Trends". Hindawi Security and Communication Networks, Vol. 2019.

[19] Fratini, A., Sansone, M., Bifulco, P. et al. Individual identification via electrocardiogram analysis. BioMed Eng OnLine 14, 78 (2015). https://doi.org/10.1186/s12938-015-0072-y

[20] - ARTEAGA-FALCONI, J. S., OSMAN, H. A., SADDIK, A. E. (2018). "ECG and Fingerprint Bimodal Authentication". Sustainable Cities and Society, Vol. 40.

[21] - TORTORA, G. J., DERRICKSON, B. (2014). "Principles of

anatomy and physiology". Livro, décima quarta edição. John Wiley & Sons.

[22] - AFONSO, V. X. (1993). "ECG QRS detection". Tompkins WJ (eds) Biomedical digital signal processing. PTR Prentice-Hall, Englewood Cliffs.

[23] - MASON, J. W., BADILINI, F., VAGLIO, M., LUX, R. L., AYSIN, B., MOON, T. E., HEINZ, B., STRACHAN, I. (2016). "A fundamental relationship between intraventricular conduction and heart rate". Journal of Electrocardiology ,Vol. 49, No. 3.

[24] - MOSENIA , A., SUR-KOLAY, S., RAGHUNATHAN, A., JHA, N. K. (2017). "CABA: Continuous authentication based on BioAura", IEEE Trans. Comput.

[25] - SOUZA, P. H. B. (2019). "Método para estimação da frequência cardíaca e variabilidade cardíaca com base em fotopletismografia por vídeo". Master's degree dissertation in Biomedical Engineering, Graduation Program in Biomedical Engineering, University of Brasília, Brasília, DF, Publication nr. 109A/2019.

[26] - KOHAVI, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection". IJCAI - International Joint Conference on Articial Intelligence.

[27] - (2000). "The MIT-BIH Arrhythmia Database". Website disponível em "http://www.physionet.org/physiobank/database/mitdb/", Accessed May 18, 2019.

[28] - (2000). "The MIT-BIH Normal Sinus Rhythm Database". Website disponível em "http://www.physionet.zorg/physiobank/database/nsrdb/", Accessed Mai 18, 2019.

[29] - TADDEI, A. (1992). "The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography". European Heart Journal, Vol. 13, No. 9.

[30] - LAGUNA, P., MARK, R. G., GOLDBERG, A., MOODY G. B. (1997). "A database for evaluation of algorithms for measurement of QT and other wave-form intervals in the ECG," Computing in Cardiology, vol. 24, pp.673-676.

[31] - BARANDELA, R., VALDOVINOS, R. M., SÁNCHEZ, J. S. (2003). "New Applications of Ensembles of Classifiers". Pattern Analysis & Applications, Vol.6(3), 245–256. https://doi.org/10.1007/s10044-003-0192-z.

[32] - ZHAO, Y., XUE, J., CHEN, X. (2014). "Ensemble Learning Approaches in Speech Recognition". In Speech and Audio Processing for Coding, Enhancement and Recognition. Pages 113 – 152. https://doi.org/10.1007/978-1-4939-1456-2_5.

[33] - PAGANO, C., GRANGER, E., SABOURIN, R., MARCIALIS, G., ROLI, F. (2014). "Adaptive ensembles for face recognition in changing video surveillance environments". Information Sciences, 286, 75–101. https://doi.org/10.1016/j.ins.2014.07.005.

[34] - Dai, P., Gwadry-Sridhar, F., Bauer, M., & Borrie, M. (2016). "Bagging Ensembles for the Diagnosis and Prognostication of Alzheimer's Disease". Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Phoenix, Arizona, USA. February 2016.

[35] - KANDALA, R., & DHULI, R. (2018). "Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier". Biomedical Signal Processing and Control, 41, 242–254. https://doi.org/10.1016/j.bspc.2017.12.004.