



UnB

UNIVERSIDADE DE BRASÍLIA – UNB

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

**Inovações para identificação automática de triatomíneos
(Hemiptera: Reduviidae) baseada no processamento de imagens
por aprendizado de máquina**

Vinícius Lima de Miranda

**Brasília-DF
Agosto/2023**



UnB

UNIVERSIDADE DE BRASÍLIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

**Inovações para identificação automática de triatomíneos
(Hemiptera: Reduviidae) baseada no processamento de imagens
por aprendizado de máquina**

Vinícius Lima de Miranda

Tese de Doutorado apresentada ao Programa de Pós-graduação em Zoologia, Instituto de Ciências Biológicas, da Universidade de Brasília, como parte dos requisitos necessários à obtenção do título de Doutor em Zoologia.

Orientador: Dr. Rodrigo Gurgel Gonçalves

Brasília
Agosto/2023

Inovações para identificação automática de triatomíneos (Hemiptera: Reduviidae)
baseada no processamento de imagens por aprendizado de máquina

Vinícius Lima de Miranda
Orientador: Dr. Rodrigo Gurgel Gonçalves

Tese de Doutorado apresentada ao Programa de Pós- graduação em Zoologia, Instituto de Ciências Biológicas, da Universidade de Brasília, como parte dos requisitos necessários à obtenção do título de Doutor em Zoologia.

Aprovada por:

Prof. Dr. Rodrigo Gurgel Gonçalves – UnB – Presidente

Prof. Dr. Antônio José Camilo de Aguiar – UnB – Membro Titular

Prof. Dr. Cléber Galvão – Fiocruz/RJ – Membro Titular

Prof. Dra. Raquel Aparecida Ferreira – Fiocruz/MG – Membro Titular

Prof. Dr. Marcos Takashi Obara – UnB – Membro Suplente

Brasília-DF
Agosto/2023

Dedico este trabalho aos meus bisavós Josefa Lima (*in memoriam*), Roberto Rodrigues (*in memoriam*) e ao meu pai Valdei Miranda (*in memoriam*).

AGRADECIMENTOS

Agradeço...

A *Deus*, pela vida, pela saúde, por me manter perseverante, por não me deixar desistir, por me guiar e por me mostrar as saídas quando tudo parecia impossível.

Ao meu orientador Dr. *Rodrigo Gurgel Gonçalves* por todo acolhimento, amizade e orientação. Agradeço por acreditar no meu potencial, por ter contribuído para que eu conseguisse me estabelecer em um lugar tão distante das minhas origens e culturas. Agradeço por sempre estar comigo em todos os momentos, bons e difíceis, nesses últimos 4 anos. Você é um grande e fantástico exemplo para mim!

A dois grandes pesquisadores e amigos, *Fernando Abad Franch* e *José Roberto Pujol*, obrigado pelos ensinamentos.

Aos membros da banca, *Dr. Antônio José Camilo de Aguiar*, *Dr. Cléber Galvão*, *Dra. Raquel Aparecida Ferreira* e o *Dr. Marcos Takashi Obara* por aceitarem contribuir com este trabalho.

Às coleções biológicas da Universidade de Brasília (UnB), do Instituto René Rachou (Fiocruz-MG), do Instituto Oswaldo Cruz (Fiocruz-RJ) e da Unesp (Araraquara-SP), por fornecer os espécimes para os estudos conduzidos nesta tese.

À minha amada e magnífica esposa, *Vanessa Abreu*, que tenho a honra de ter ao meu lado partilhando desta conquista juntos. Agradeço por ter me aceitado em um dos momentos mais desafiadores da minha vida. Você é sensacional, eu te amo muito!

À minha Mãe, *Patrícia Lima*, por sempre manter o entusiasmo, o amor e a alegria que contagia e que apesar da distância, sempre se mostra presente cativando a todos.

Aos meus irmãos *Jhenifer Lima* e *Valter Alves*, por todo o apoio, motivação e compreensão nos momentos em que mais precisei.

Aos meus tios *Maria Nazaré* e *Abílio Oliveira* por todo amor e assistência que me deram desde a minha infância, obrigado por tudo.

Aos meus cunhados, *Alcileide Medrade* e *Vinícius Abreu*, por todo o apoio.

À minha orientadora da graduação *Dra. Maria Regiane Araujo Soares*, a precursora para que acontecesse essa minha trajetória acadêmica. Você é uma mulher fantástica!

Aos meus orientadores do mestrado, *Dra. Cléa Mariano* e *Dr. Jacques Delabie*, que contribuíram muito para o meu amadurecimento na ciência. Meu grande obrigado!

Aos estudantes de iniciação científica, mestrado e doutorado do Laboratório de Parasitologia Médica e Biologia de Vetores da Universidade de Brasília (UnB).

Aos meus amigos do Piauí, Bahia e Brasília, *Luciano Neto*, *Mário Júnior*, *Mauro Aparecido*, *Braion Moura*, *Breendo Moura*, *Priscilla Silva*, *Jéssica Fernandes*, *Cristiele Souza*, *Mayla Silva*, *Lucas Carneiro*, *Fabício Leal*, *Taís Araújo* e *Patrícia Gomes*.

Aos meus sogros, *Maria Abreu* e *Sérgio Abreu*, pessoas maravilhosas que entraram na minha vida.

A todo o corpo docente do curso de graduação em Ciências Biológicas da Universidade Federal do Piauí (UFPI/CAFS), do Programa de Pós-graduação em Zoologia da Universidade Estadual de Santa Cruz (PPGZOO/UESC) e do Programa de Pós-graduação em Zoologia da Universidade de Brasília (PPGZOO/UnB).

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro para condução do projeto de pesquisa. Ao auxílio PROAP/CAPES por todo apoio financeiro que contribuiu para o desenvolvimento deste trabalho.

SUMÁRIO

RESUMO.....	9
ABSTRACT.....	12
INTRODUÇÃO GERAL.....	15
Triatomíneos	15
Biologia e comportamento.....	18
Importância em saúde pública	21
Identificação de triatomíneos	22
Aprendizado de máquina	25
Métodos de identificação automática de insetos	26
Identificação automática de triatomíneos	30
JUSTIFICATIVA.....	32
OBJETIVOS.....	34
Geral	34
Específicos.....	34
REFERÊNCIAS.....	35
CAPÍTULO I.....	46
Cellphone picture-based, genus-level automated identification of Chagas disease vectors: effects of picture orientation on the performance of five machine-learning algorithms	46
Abstract	50
1. Introduction.....	52
2. Methods	54
3. Results.....	63
4. Discussion.....	71
5. Conclusions and outlook	75
References	77
CAPÍTULO II.....	154
Identificação automática de imagens dorsais de triatomíneos adultos ao nível de espécie usando a rede neural convolucional pré-treinada AlexNet	154
Resumo.....	155
1. Introdução	155
2. Material e métodos	157
3. Resultados	160
4. Discussão.....	171
5. Conclusão	177
6. Referências	178
CAPÍTULO III.....	193
Identificação automática de ninfas de triatomíneos ao nível de gênero usando a rede neural convolucional pré-treinada AlexNet	193
Resumo.....	194

1. Introdução	194
2. Material e métodos	196
3. Resultados	200
4. Discussão	209
5. Conclusão	211
6. Referências	213
CONSIDERAÇÕES FINAIS	222

RESUMO

Inovações para identificação automática de triatomíneos (Hemiptera: Reduviidae) baseada no processamento de imagens por aprendizado de máquina

Vinícius Lima de Miranda

Orientador: Dr. Rodrigo Gurgel Gonçalves

Resumo da Tese de Doutorado apresentada ao Programa de Pós-graduação em Zoologia, Instituto de Ciências Biológicas, da Universidade de Brasília, como parte dos requisitos necessários à obtenção do título de Doutor em Zoologia.

Os triatomíneos são insetos hemípteros (Reduviidae), transmissores de *Trypanosoma cruzi*, agente etiológico da doença de Chagas. A identificação automática de triatomíneos pode fortalecer a vigilância e o controle da transmissão vetorial da doença de Chagas. Essa tese apresenta três capítulos com inovações para identificação automática de triatomíneos. No primeiro capítulo, avaliamos o desempenho de cinco algoritmos de aprendizado de máquina para identificação de três gêneros de triatomíneos com base em imagens em diferentes ângulos ou posições obtidas com uma câmera de celular. Analisamos imagens de 730 insetos (13 espécies) fotografados em nove ângulos representando três posições. Dividimos o banco de dados de 6570 imagens em conjuntos de treinamento (80%) e de teste (20%) e, em seguida, treinamos e testamos uma rede neural convolucional pré-treinada (AlexNet, AN), três classificadores (AdaBoost, AB; Gradient Boosting, GB; Histogram-based Gradient Boosting, HB) e um modelo discriminante linear (LD). Avaliamos a acurácia e a especificidade da identificação com modelos generalizados lineares mistos. Os resultados mostraram que as diferenças de desempenho entre os algoritmos foram impulsionadas pela acurácia e especificidade essencialmente perfeitas da AN, independentemente do ângulo ou posição. As acurácias previstas de HB variaram de ~0,987 (*Panstrongylus*) a >0,999 (*Triatoma*). A acurácia do AB foi ruim para *Rhodnius* (~0,224-0,282) e *Panstrongylus* (~0,664-0,729), mas razoável para *Triatoma* (~0,988-0,991). Para *Panstrongylus*, LD e GB previram acurácias na faixa de ~0,970-0,984. AB classificou erroneamente ~57% de *Rhodnius* e *Panstrongylus* como *Triatoma*, enquanto as estimativas de especificidade variaram de ~0,92 a ~1,0 para as demais combinações de gênero-algoritmo. No segundo capítulo, avaliamos o

desempenho da rede AN na identificação de espécies de triatomíneos a partir de imagens dorsais de alta resolução e obtidas de celulares comuns. Analisamos fotos dorsais de 6397 triatomíneos pertencentes a 65 espécies de sete gêneros, a partir de um conjunto de dados de treinamento (70%) e teste (20%). Os 10% restantes foram utilizados para a validação interna do algoritmo durante o treinamento, realizado no MATLAB. Realizamos dois experimentos, um com todas as espécies e outro com as espécies com registro em casas e infecção natural por *T. cruzi*. Os resultados mostraram que AN apresentou uma acurácia de ~0,93 (IC95%, 0,913-0,941) para identificar 65 espécies de triatomíneos. A acurácia específica alcançou valor máximo para 21 espécies, principalmente as espécies dos gêneros *Rhodnius* e *Panstrongylus*. O desempenho foi melhorado (~0,97, IC95%: 0,930-0,958) quando foram consideradas apenas as espécies com maior capacidade vetorial. No terceiro capítulo, avaliamos o desempenho da AN na identificação de imagens de ninfas de triatomíneos até o nível de gênero. A abordagem metodológica foi similar ao capítulo 2, porém foram realizados três experimentos. No primeiro, treinamos a rede AN para diferenciar ninfas de triatomíneos e espécimes não-triatomíneos (adultos e ninfas de percevejos heterópteros). No segundo, a tarefa foi diferenciar ninfas de triatomíneos de ninfas de heterópteros não-triatomíneos. No terceiro, a tarefa foi diferenciar ninfas dos três gêneros de triatomíneos mais relevantes em saúde pública, *Panstrongylus*, *Rhodnius* e *Triatoma*. Os resultados mostraram que a acurácia média para identificação de ninfas de triatomíneos x outros heterópteros (ninfas e/ou adultos) variou de 0,97 a 1,00. Os valores de acurácia também foram altos para identificação de ninfas dos gêneros *Triatoma* (~0,99, IC95%: 0,947-0,998) e *Rhodnius* (~1,00, IC95%: 0,816-1,000), sendo menores para *Panstrongylus* (~0,58, IC95%: 0,360-0,784). As seguintes conclusões derivam dessa tese: 1) quando algoritmos de aprendizado de máquina como AN são usados, os ângulos ou posições em que os insetos são fotografados não são relevantes para identificar corretamente imagens de celulares de triatomíneos em nível de gênero, com acurácia e especificidade essencialmente perfeitas. 2) A rede AN apresentou ótimo desempenho para identificação automática de espécies de triatomíneos baseada em um conjunto amplo de imagens dorsais de adultos. 3) AN apresentou ótimo desempenho para identificar ninfas de triatomíneos, diferenciando de ninfas e adultos de heterópteros em geral, ninfas de triatomíneos e não triatomíneos e ninfas de *Rhodnius* e *Triatoma*. Finalmente, a tese fornece a base para o desenvolvimento de um sistema

automatizado de identificação dos vetores da doença de Chagas com participação comunitária em larga escala, o que poderá contribuir para a ciência cidadã.

Palavras-Chave. Triatominae; aprendizado de máquina; câmeras de celulares; posição da imagem; ninfas; vigilância entomológica; ciência cidadã.

ABSTRACT

Innovations for automatic identification of triatomines (Hemiptera: Reduviidae) based on machine learning image processing

Vinícius Lima de Miranda

Orientador: Dr. Rodrigo Gurgel Gonçalves

Abstract da Tese de Doutorado apresentada ao Programa de Pós-graduação em Zoologia, Instituto de Ciências Biológicas, da Universidade de Brasília, como parte dos requisitos necessários à obtenção do título de Doutor em Zoologia.

Triatomines bugs (Reduviidae) are vectors of *Trypanosoma cruzi*, the etiologic agent of Chagas disease. Automated identification of triatomine bugs could strengthen the control/surveillance of vector-borne transmission of Chagas disease. This thesis presents three chapters with innovations for automatic identification of triatomines. In the first chapter, we assess the performance of five machine-learning algorithms at identifying the three main triatomine genera based on bugs pictured at different angles/positions with an ordinary cellphone camera. Each of 730 bugs (13 species) was pictured at nine angles representing three positions. We randomly split the 6570-picture database into training (80%) and testing sets (20%), and then trained and tested a pre-trained convolutional neural network (AlexNet, AN); three boosting-based classifiers (AdaBoost, AB; Gradient Boosting, GB; and Histogram-based Gradient Boosting, HB); and a linear discriminant model (LD). We assessed identification accuracy and specificity with generalized linear mixed models. Differences in performance across algorithms were mainly driven by AN's essentially perfect accuracy and specificity, irrespective of picture angle or bug position. HB predicted accuracies ranged from ~0.987 (*Panstrongylus*) to >0.999 (*Triatoma*). AB accuracy was poor for *Rhodnius* (~0.224-0.282) and *Panstrongylus* (~0.664-0.729), but fair for *Triatoma* (~0.988-0.991). For *Panstrongylus*, LD and GB had predicted accuracies in the ~0.970-0.984 range. AB misclassified ~57% of *Rhodnius* and *Panstrongylus* as *Triatoma*, whereas specificity estimates ranged from ~0.92 to ~1.0 for the remaining algorithm-genus combinations. In the second chapter, we evaluated the performance of the AN network in identifying triatomine species based on dorsal pictures obtained from ordinary cellphone camera and high-quality images. We analyzed dorsal pictures

of 6397 triatomines belonging to 65 species of seven genera, from a training (70%) and test (20%) dataset. The remaining 10% dataset were used for the internal validation of AN network during training in MATLAB. We carried out two experiments, one with all species and the other with species registered in houses and naturally infected by *T. cruzi*. AN presented an accuracy of ~0.93 (95%CI, 0.913-0.941) to identify 65 triatomine species. The specific accuracy reached a maximum value for 21 species, mainly *Rhodnius* and *Panstrongylus* species. Performance was improved (~0.97, 95%CI: 0.930-0.958) in the analysis including only species with greater vectorial capacity. In the third chapter, we evaluated the performance of AN in identifying images of triatomine nymphs. The methodological approach was similar to Chapter 2, but three experiments were performed. In the first, we train the AN network to differentiate between triatomine nymphs and non-triatomine specimens (adults and nymphs of heteropterans). In the second, the task was to differentiate triatomine nymphs from non-triatomine nymphs. In the third, the task was to differentiate nymphs from the three most relevant triatomine genera in public health, *Panstrongylus*, *Rhodnius* and *Triatoma*. The results showed that the mean accuracy for identifying triatomine nymphs x other heteropterans (nymphs and/or adults) ranged from 0.97 to 1.00. Accuracy values were also high for identifying nymphs of the genera *Triatoma* (~0.99, 95%CI: 0.947-0.998) and *Rhodnius* (~1.00, 95%CI: 0.816-1.000), being lower for *Panstrongylus* (~0.58, 95%CI: 0.360-0.784). The following conclusions derive from this thesis: 1) when machine learning algorithms such as AN are used, the angles or positions at which insects are photographed are not relevant to correctly identify cellphone images of triatomine bugs at the genus level, with accuracy and specificity essentially perfect. 2) the AN network showed excellent performance for the automatic identification of species based on a large set of dorsal images of adult triatomines. 3) AN showed excellent performance in identifying triatomine nymphs from other heteropterans (adults or nymphs), triatomine and non-triatomine nymphs, and nymphs of *Rhodnius* and *Triatoma*. Finally, the thesis provides the basis for the development of an automated system for identifying Chagas disease vectors with large-scale community participation, which could contribute to citizen science.

Key words. Triatominae; machine learning; cell phone cameras; image position; nymphs; entomological surveillance; citizen science.

Brasília
Agosto/2023

INTRODUÇÃO GERAL

Triatomíneos

Hemiptera Linnaeus, 1758 é a ordem mais diversa dentre os insetos hemimetábolos, sendo conhecidas mais de 90.000 espécies, divididas em três subordens: Sternorrhyncha Amyot & Audinet-Serville, 1843, Auchenorrhyncha Carl Stål, 1866 e Heteroptera Latreille, 1810 (Gullan & Cook, 2007; Panizzi & Grazia, 2015; Bartlett *et al.*, 2018). São caracterizados pela presença de aparelho bucal do tipo picador-sugador, com canal de sucção formado pela justaposição dos estiletes maxilares envolvidos pelos estiletes mandibulares, além de serem providos de dois pares de asas, sendo o par anterior mais rígido do que o par posterior (Gallo *et al.*, 2002). A maioria das espécies de hemípteros são fitófagas, com exceção dos heterópteros da família Cimicidae e Reduviidae, compostos, em sua maioria, por organismos com hábitos alimentares predatórios e algumas espécies hematófagas (Goodchild, 1966; Cohen *et al.*, 1990).

Os triatomíneos, popularmente conhecidos como barbeiros no Brasil, são insetos hematófagos pertencentes à família Reduviidae, que atuam como vetores de *Trypanosoma cruzi* Chagas, 1909, agente etiológico da doença de Chagas (Lent & Wygodzinsky, 1979). A maioria das mais de 150 espécies de triatomíneos conhecidas ocorrem nas Américas (Monteiro *et al.*, 2018, Zhao *et al.*, 2023), para as quais as tradicionais pesquisas são relativas à sua importância na transmissão do agente causador da doença de Chagas (Noireau & Dujardin, 2010).

Os triatomíneos evoluíram de insetos reduvídeos predadores, muito provavelmente no Novo Mundo, e se diversificaram pelas Américas. Decorrente da grande quantidade de estudos já realizados, tendo como modelo biológico os Reduviidae Latreille, 1807, catálogos modernos, livros de referência e chaves de identificação podem divergir no número de subfamílias deste grupo (Maldonado Capriles, 1990; Schuh & Slater, 1995; Weirauch *et al.*, 2014). Foram classificados pela primeira vez como uma tribo por Jeannel (1919) e em seguida elevado ao status de família por Pinto (1926), vindo a ser tratados como uma subfamília de Reduviidae predadores com Usinger (1943). No entanto, apesar da falta de consenso sobre qual o grupo de predador deu origem aos triatomíneos, Reduviinae Latreille, 1807 seria a provável subfamília de Reduviidae filogeneticamente mais próxima de Triatominae Jeannel, 1919 (Weirauch, 2008; Monteiro *et al.*, 2018), porém o esclarecimento dessa

questão é dificultado pela escassez de registro fóssil (Galvão, 2003; Lukashevich & Mostovski, 2003).

A revisão mais atualizada sobre a sistemática dos Triatominae foi proposta por Monteiro *et al.* (2018), que apresentaram uma visão geral da evolução, sistemática e biogeografia desses vetores, discutindo a partir de suas compreensões atuais, quantas vezes e quando a hematofagia (autapomorfia que define a subfamília) poderia ter evoluído. Além disso, essa revisão apresenta um resumo dos recentes avanços na sistemática deste grupo de insetos, com ênfase na contribuição de ferramentas moleculares para o esclarecimento de controvérsias taxonômicas, sendo também uma síntese global da biogeografia dos triatomíneos (Monteiro *et al.*, 2018).

Apesar da grande quantidade de contribuições em estudos morfológicos com triatomíneos utilizando ferramentas como a morfometria linear e geométrica e baseando na comparação de caracteres morfológicos (Jurberg & Galvão, 1997; Dujardin *et al.*, 1999; Monroy *et al.*, 2003), ainda é necessária a utilização de métodos moleculares, que na maioria dos casos é fundamental para explicar as diversas hipóteses sobre a história evolutiva deste grupo (Lehmann *et al.*, 2005; Mendonça *et al.*, 2009). Desde o início do século XXI, as análises filogenéticas baseadas em morfologia e com amostragem de táxons e caracteres muito diferentes têm como objetivo esclarecer as relações entre Reduviidae predadores e hematófagos. A Figura 1 apresenta o resumo das hipóteses filogenéticas de Reduviidae publicadas, apresentando as relações propostas entre Triatominae e Reduviidae predadores específicos (Monteiro *et al.*, 2018).

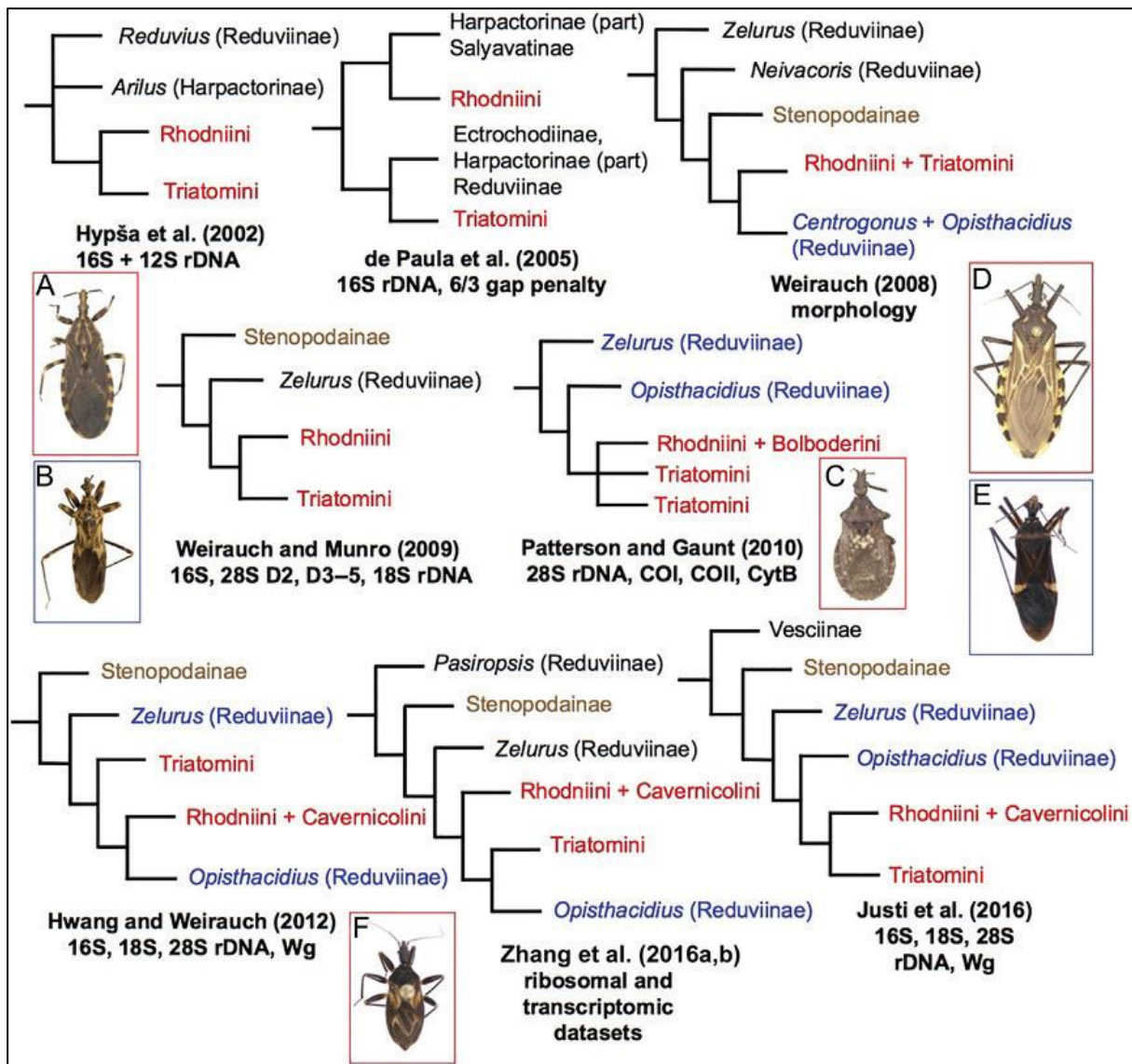


Figura 1. Hipóteses filogenéticas de Reduviidae já publicadas, mostrando os triatomíneos que se alimentam de sangue, táxon pertencente ao clado *Zelurus* Hahn, 1826 dos Reduviinae polifiléticos, Stenopodainae Amyot & Serville, 1843 e outros Reduviidae. (A) *Rhodnius paraensis* Sherlock, Guitton & Miles, 1977 (Triatominae: Rhodniini); (B) *Opisthacidius picturatus* Lent & Wygodzinsky, 1956 (Reduviinae); (C) *Microtriatoma borbai* Lent & Wygodzinsky, 1979 (Triatominae: Bolboderini); (D) *Triatoma melanocephala* Neiva & Pinto, 1923 (Triatominae: Triatomini); (E) *Zelurus* sp. (Reduviinae); e (F) *Cavernicola pilosa* Barber, 1937 (Triatominae: Cavernicolini). Fonte: Monteiro *et al.*, 2018.

As primeiras classificações se baseavam no grau de similaridade e diferença morfológica observado entre os organismos, sem base em qualquer estrutura filogenética. No entanto, espécies morfológicamente mais semelhantes não estão necessariamente mais relacionadas. A taxonomia tradicional baseada na morfologia

evoluiu recentemente para incorporar novos tipos de dados, como por exemplo, moleculares, ecológicos, biogeográficos ou comportamentais, originando a taxonomia integrativa, que ainda tem sido pouco utilizada na taxonomia de triatomíneos (Dayrat, 2005; Monteiro *et al.*, 2018; Alevi *et al.*, 2021). Nas análises mais recentes baseadas em diferentes marcadores moleculares é demonstrado que reduvídeos predadores dos gêneros *Zelurus* e *Opisthacidius* Berg, 1879 estariam mais relacionados com os triatomíneos (Figura 1). Monteiro *et al.* (2018) apresenta, além de dados morfológicos (referência a outros estudos) e moleculares, dados biogeográficos, considerando a distribuição geográfica dos triatomíneos a partir de dados de ocorrência.

Biologia e comportamento

Os triatomíneos são insetos com desenvolvimento hemimetábolo. Após reprodução sexuada, as fêmeas produzem ovos, os quais eclodem gerando ninfas que passam por cinco estádios (Figura 2) até chegar a fase adulta. A necessidade de hematofagia durante o desenvolvimento os levou a estabelecer relações ecológicas com os seus hospedeiros vertebrados, incluindo humanos. Ao contrário de outros insetos sugadores de sangue, tais como mosquitos e flebotomíneos, nos quais apenas a fêmea é hematófaga, os dois sexos e todas as fases ninfaís dos triatomíneos alimentam-se de sangue. Algumas espécies defecam durante ou logo após a alimentação, sendo consideradas transmissoras com maior capacidade vetorial de *Trypanosoma cruzi* (Lazzari, 2014).

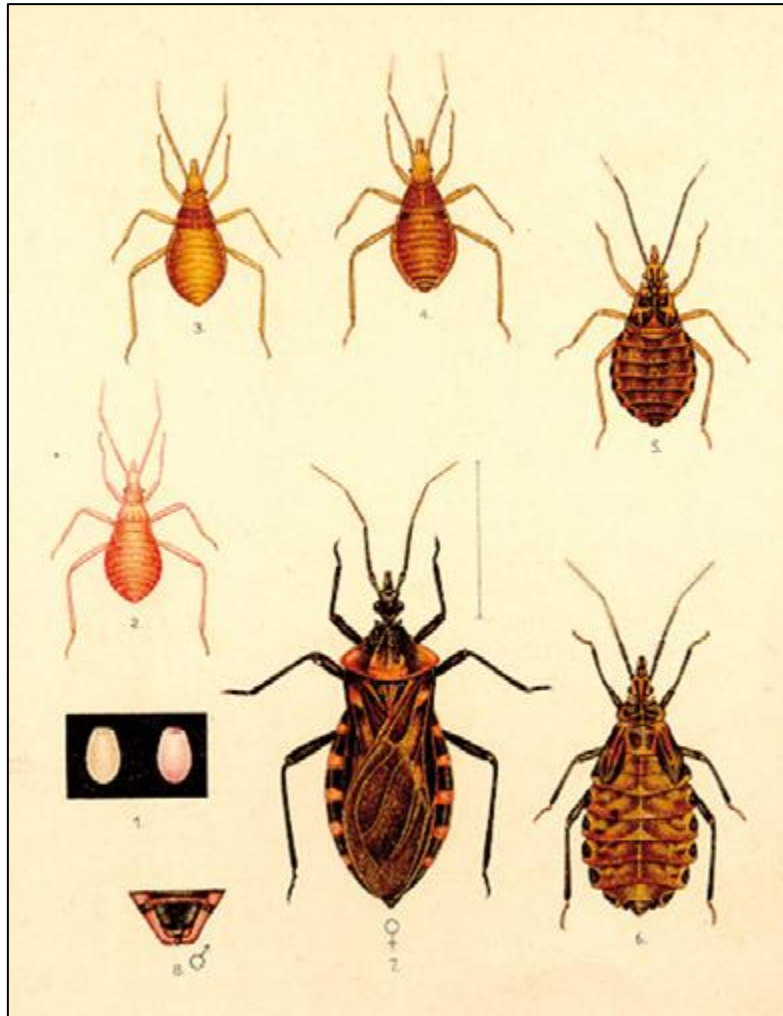


Figura 2. Representação do ciclo de desenvolvimento dos triatomíneos, desde ovos, estádios ninfais, e adultos com a diferenciação da terminália de um macho e uma fêmea de *Panstrongylus megistus* Burmeister, 1835. Fonte: Jurberg *et al.*, 2005 (ilustração elaborada por Raymundo Honório).

As diversas formas pelas quais os triatomíneos se relacionam com seus microhabitats e hospedeiros podem ser organizadas em dois principais comportamentos. As espécies de triatomíneos especializadas em ninhos se comportam em grande parte como forrageadoras 'senta-e-espera' associadas aos ninhos (incluindo tocas, covas, abrigos). Esses triatomíneos vivem e se alimentam nos ninhos de seus hospedeiros, da mesma maneira que os cimicídeos. Por outro lado, espécies de insetos oportunistas-generalistas de forrageamento ativo ocupam e se reproduzem em microhabitats como afloramento rochosos, troncos de árvore ou copas de palmeiras independentemente de suas fontes alimentares nidificarem lá. Esses triatomíneos se alimentam dos hospedeiros que ocupam esses microhabitats e

se comportam em grande parte como micropredadores que localizam, perseguem e realizam hematofagia nos animais presentes (Abad-Franch & Gurgel-Gonçalves, 2021). Esses estilos de vida são evidenciados por adaptações morfológicas que podem ser percebidas nos grupos de triatomíneos. Por exemplo, especialistas em ninhos tendem a ser menores do que os triatomíneos 'perseguidores'. As cabeças dos insetos especialistas em ninhos também são particularmente curtas e fortes (Figura 3). Além disso, 'Perseguidores' tendem a ser mais agressivos e mais rápidos ao se aproximar das fontes alimentares do que os especialistas em ninhos (Abad-Franch & Gurgel-Gonçalves, 2021).

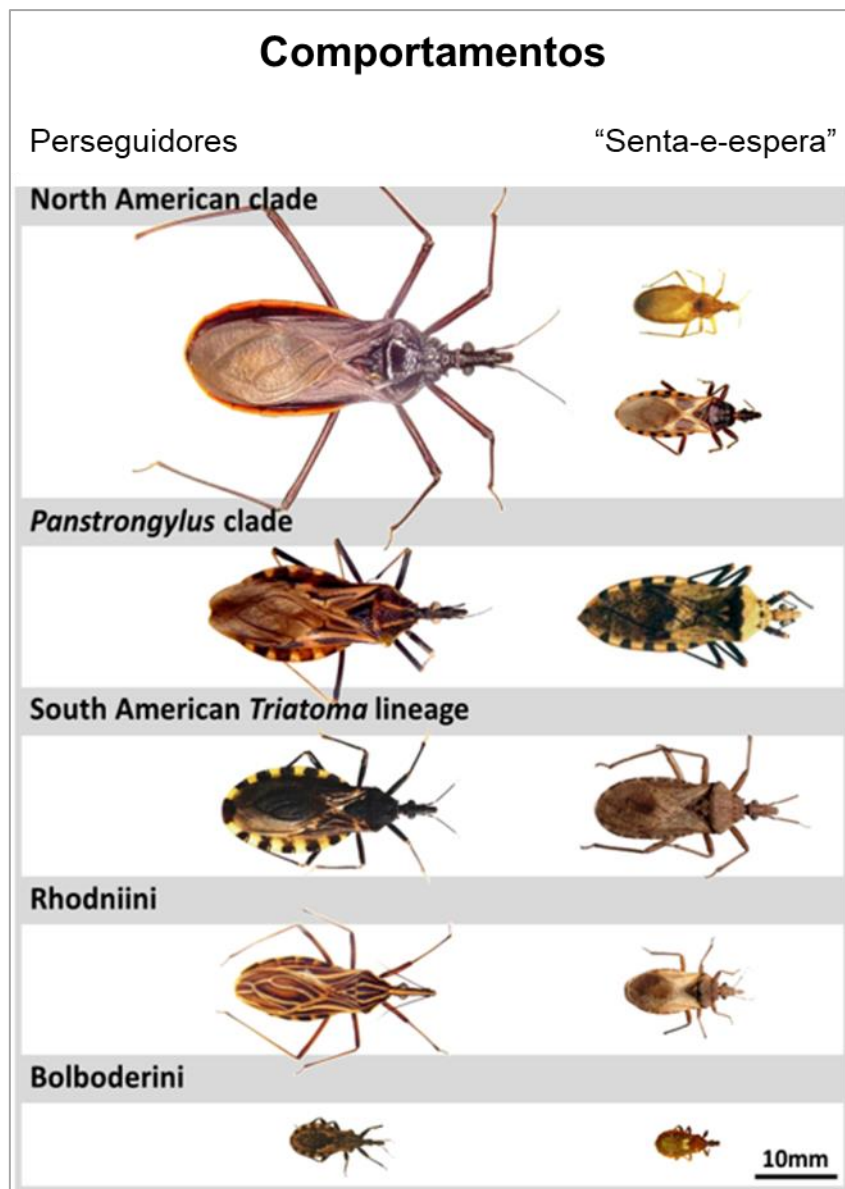


Figura 3. Exemplos de triatomíneos que apresentam diferentes comportamentos: 'perseguidores' e 'senta-e-espera ou associadas a ninhos. Fonte: adaptado de Abad-Franch & Gurgel-Gonçalves, 2021.

Importância em saúde pública

Os triatomíneos podem transmitir o agente etiológico *T. cruzi*, um protozoário flagelado da família Trypanosomatidae Doflein, 1901 (Cavalier-Smith *et al.*, 2016) o qual já foi registrado em aproximadamente 200 espécies de mamíferos de sete ordens, no continente americano (Roque & Jansen, 2015). Após a infecção natural, os triatomíneos podem permanecer infectados por toda a vida (Sherlock *et al.*, 1997). Na forma clássica da transmissão, a infecção em humanos é causada quando o triatomíneo defeca na pele durante a hematofagia, liberando assim formas tripomastigotas metacíclicas de *T. cruzi*. Dessa forma, o parasito penetra em pequenas lesões da pele ou em mucosas, onde se desenvolvem dentro de macrófagos e depois se distribuem pelo corpo, multiplicando-se e ocasionando lesões por células musculares cardíacas e lisas, e em casos extremos, insuficiência cardíaca, podendo levar os infectados a óbito, e fazendo com que a doença de Chagas seja classificada como a quarta enfermidade de maior impacto social na América Latina, quando comparada a todas as doenças infecciosas e parasitárias (Schmunis, 2000; Galvão, 2014).

A Organização Mundial da Saúde (OMS) estima que 6 a 7 milhões de pessoas são afetadas pela doença de Chagas em todo o mundo, principalmente na América Latina (WHO, 2019). Graças a campanhas efetivas de controle de vetores de *T. cruzi*, a infestação das casas pelas espécies de triatomíneos em alguns países da América do Sul [e.g. *T. infestans* (Klug, 1834) no Brasil] (Dias, 2007) e da América Central (e.g. *R. prolixus* Stål, 1859 na Guatemala) (Hashimoto & Schofield, 2012) quase foi eliminada. Porém, as estimativas de incidência da OMS (~40.000 novas infecções por ano) (WHO, 2019), sugerem que a transmissão ainda é relevante, sendo que a maioria dos novos casos é ocasionada por vetores nativos, que permanecem invadindo e colonizando as residências nas Américas (Abad-Franch *et al.*, 2013; Abad-Franch, 2016, Rojas de Arias *et al.*, 2022).

Um cenário epidemiológico em que a transmissão é mediada por diferentes espécies de vetores nativos com populações peridomiciliares e eventualmente domiciliares exige uma vigilância entomológica mais ampla, não apenas concentrada em uma espécie de vetor "primária" (e.g. *T. infestans*) e algumas "secundárias" [e.g. *T. brasiliensis* Neiva, 1911, *T. sordida* (Stål, 1859) e *T. pseudomaculata* Correa & Espínola, 1964] (Silveira & Martins, 2014; Waleckx *et al.*, 2015). Nesse novo cenário

há necessidade de ampliar a vigilância para outras espécies nativas. Por exemplo, espécimes de *Panstrongylus tibiamaculatus* (Pinto, 1926) infectados têm invadido casas Bahia (Ribeiro Jr *et al.*, 2015) e o mesmo ocorre com outras espécies como *R. pictipes* Stål, 1872, *R. robustus* Larrousse 1927, *R. neglectus* Lent 1954, *T. costalimai* Verano & Galvão, 1958, *P. geniculatus* (Latreille, 1811) em Tocantins (Brito *et al.*, 2017a, 2017b). Dessa forma, o serviço de vigilância entomológica da doença de Chagas deve estar preparado para identificar uma maior quantidade de espécies que podem colonizar ou invadir o intra e o peridomicílio, o que exigirá estratégias de identificação eficazes.

Identificação de triatomíneos

Considerando o grande número de espécies com diferentes comportamentos eco-epidemiológicos, a correta identificação taxonômica conduzida por profissionais treinados desempenha um papel importante no desenvolvimento e aprimoramento de estratégias de controle e vigilância de vetores (Oliveira *et al.*, 2017). Atualmente, as chaves dicotômicas são o principal método para identificação de triatomíneos. No entanto, o uso de chaves dicotômicas requer treinamento e conhecimento da morfologia do grupo biológico que é trabalhado (Oliveira *et al.*, 2017).

As chaves de identificação de triatomíneos mais abrangentes foram publicadas por Lent & Wygodzinsky (1979) (Figura 4). Nesse estudo há detalhamento das estruturas externas e da terminologia nos Triatominae o que facilita a compreensão de chaves. O trabalho também apresenta as descrições de todas as espécies até 1979 e as chaves dicotômicas em três idiomas. Posteriormente foram publicados Atlas dos Vetores da doença de Chagas nas Américas incluindo chaves pictóricas das espécies até 1997 (Carcavallo *et al.*, 1997). Uma atualização dessa chave pictórica foi realizada para as espécies que ocorrem no Brasil (Figura 5) (Galvão & Dale, 2014).

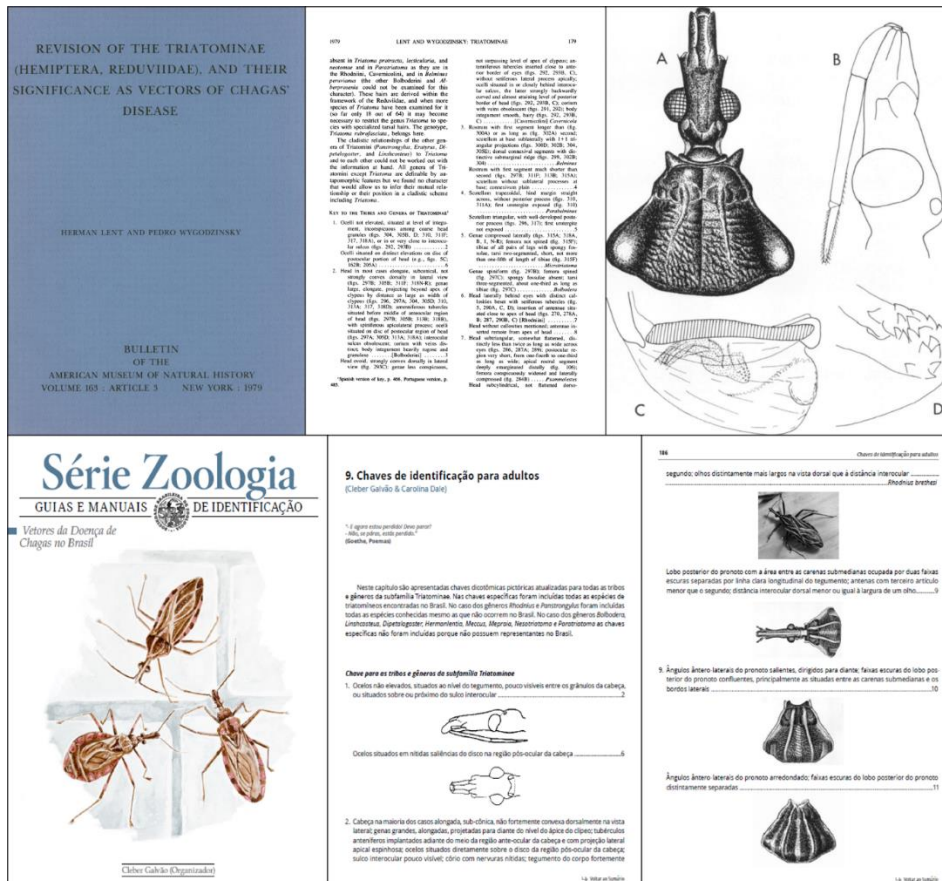


Figura 4. Principais chaves de identificação taxonômica para os triatomíneos do Brasil. Fonte: Lent & Wygodzinsky, 1979; Galvão & Dale, 2014.

Uma alternativa ao uso das chaves dicotômicas e pictóricas tradicionais descritas acima é o desenvolvimento de métodos de identificação interativos baseados em dicotomias e politomias, como é o caso da plataforma online “Xper”, que pode carregar uma grande quantidade de informações sobre diversos táxons, considerando caracteres morfológicos e ilustrações dos indivíduos (Ung *et al.*, 2010). As chaves impressas também têm sido substituídas por chaves eletrônicas (Walter & Winterton, 2007), que são mais fáceis de atualizar e podem ser incluídas em smartphones. Nesse contexto, Oliveira *et al.* (2017) desenvolveram uma ferramenta web, chamada “TriatoKey”, que utiliza uma chave pictórica e dicotômica eletrônica para identificação de 42 espécies de triatomíneos registradas no Brasil. A plataforma adota uma série de perguntas do tipo “sim” ou “não” permitindo ao usuário ser conduzido na identificação do táxon. Essa identificação é baseada em caracteres morfológicos visualizados por fotografias de espécies catalogadas da Coleção de Vetores de Tripanosomatídeos (Fiocruz/COLVEC), havendo também a possibilidade de enviar fotos de insetos para que especialistas possam confirmar ou realizar a

identificação (Souza *et al.*, 2014; Oliveira *et al.*, 2017). Outros recursos interessantes elaborados para identificação de espécies de importância médica e veterinária, são os aplicativos para celulares com sistema operacional Android, como o LutzDex™ (Rocha *et al.*, 2020) e o TriatoDex (Gurgel-Gonçalves *et al.*, 2021), que se baseiam na implementação de chaves de identificação práticas e acessíveis, a partir da organização de um banco de dados com caracteres morfológicos para espécies de flebotomíneos e triatomíneos, respectivamente.

O aplicativo TriatoDex inclui uma chave eletrônica composta por perguntas sobre características morfológicas das 150 espécies existentes até 2017. O usuário ao selecionar uma opção será levado diretamente a outra pergunta até finalizar a identificação, quando se apresenta a foto colorida do vetor identificado, sua distribuição geográfica e outras informações (habitat, importância médica, autores e ano de descrição). Para cada resposta existem imagens das estruturas para facilitar tomada de decisão do usuário durante a identificação. Em qualquer momento da identificação, o TriatoDex mostra as possíveis espécies e ainda permite retornar para perguntas anteriores a critério do usuário (Figura 5). O aplicativo apresenta uma opção para o usuário analisar todos os passos utilizados e verificar possíveis erros durante a identificação. Os erros podem ser excluídos, reiniciando o processo a partir da última pergunta correta respondida. Durante ou após a identificação final é possível redefinir a pesquisa e reiniciar uma nova tentativa. A validação do aplicativo (Gurgel-Gonçalves *et al.* 2021) mostrou que a identificação baseada no TriatoDex estava correta em 78,9% das 824 tarefas de identificação dos participantes. Como esperado, o TriatoDex claramente teve um desempenho melhor nas mãos de especialistas treinados (90,3% vs. 68,6% de identificações corretas). Em uma comparação piloto com uma chave pictórica impressa o TriatoDex teve desempenho semelhante (~84,5% das atribuições corretas, IC 68,9-94,0%), mas a identificação foi ~33% (~2,3 minutos) mais rápida, em média.

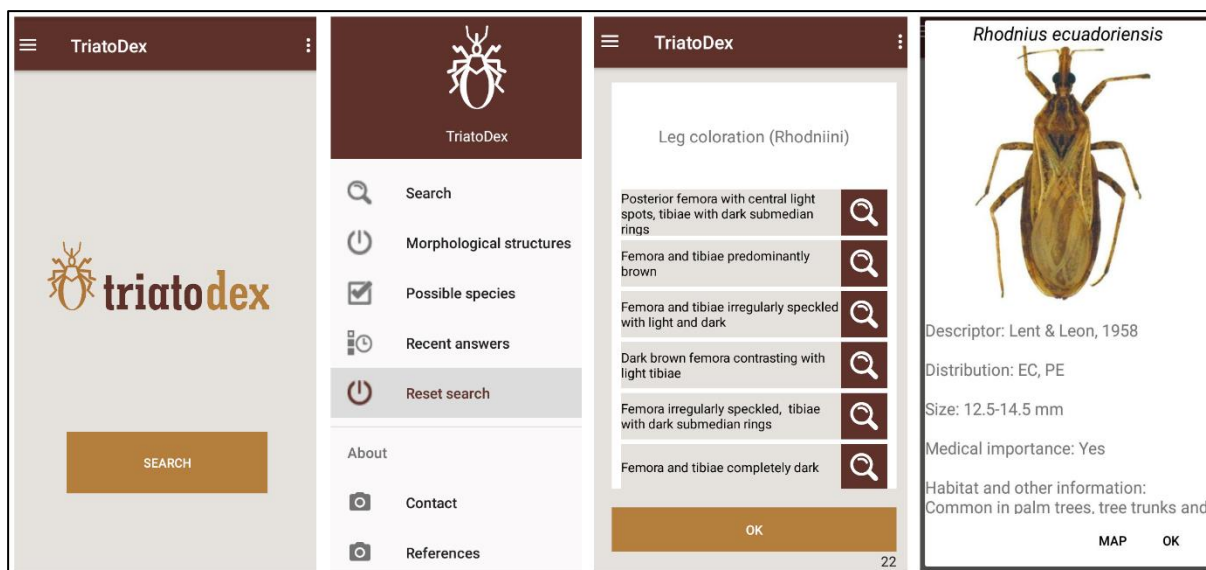


Figura 5. Interface do aplicativo de identificação taxonômica de triatomíneos, TriatoDex. Fonte: Gurgel-Gonçalves *et al.*, 2021.

Embora essas novas ferramentas facilitem a identificação, ainda necessitam de pessoas com treinamento adequado para detectar e diferenciar as características morfológicas apresentadas. Uma solução promissora é o desenvolvimento de identificação automatizada de espécies de triatomíneos baseada no processamento e análise de imagens por aprendizado de máquina.

Aprendizado de máquina

O aprendizado de máquina é um dos principais ramos da inteligência artificial, pois permite a extração de padrões significativos a partir de exemplos, sendo semelhante a inteligência humana (Erickson, *et al.*, 2017). Está representado por um conjunto de métodos, técnicas e ferramentas que auxiliam na resolução de problemas, em diversas áreas da ciência. Por exemplo, esses métodos vêm sendo usados para análises dos parâmetros clínicos, resultando na previsão da progressão de doenças, aplicação de conhecimentos médicos para avaliação de resultados, planejamento e suporte de terapia para gerenciamento de pacientes (Magoulas & Prentza, 1999; Kononenko, 2001). Apesar de ser relativamente recente, o conhecimento sobre o aprendizado de máquina aplicado ao reconhecimento de imagens médicas já está sendo utilizado há décadas, principalmente nas áreas de diagnóstico auxiliado por computadores e mapeamento cerebral funcional (Wernick *et al.*, 2010). Além disso, o aprendizado de máquina também pode ser utilizado para análise de dados,

interpretando adequadamente dados imperfeitos. Sendo assim, a correta implementação dos métodos pode integralizar sistemas computacionais em ambientes de assistência médica, facilitando e aprimorando a qualidade do trabalho na área médica (Magoulas & Prentza, 1999).

A princípio, os algoritmos de aprendizado de máquina foram desenvolvidos para analisar conjuntos de dados médicos, e atualmente essa técnica fornece diversas ferramentas indispensáveis, que se apresentam como uma tecnologia adequada para análise inteligente de dados (Kononenko, 2001). O aprendizado de máquina é subclassificado em duas categorias: 1) “aprendizado supervisionado” que tem o objetivo de prever um resultado ou um objetivo conhecido, como por exemplo, identificar imagens de objetos (se é um carro ou um avião) ou identificação de documentos (trata-se de um ensaio clínico sobre insuficiência cardíaca ou um relatório financeiro), envolvendo a escolha entre subgrupos, objetivando descrever uma nova instância de dados e previsão; 2) “aprendizado não supervisionado”, onde não há resultados previstos, sendo baseado na tentativa de encontrar padrões ou agrupamentos que ocorrem naturalmente nos dados, no qual um conjunto de dados de aprendizagem (rotulados) são usados para treinar um sistema e com isso seja possível identificar um conjunto de dados para teste (não rotulado e não utilizados no treinamento) (Deo, 2015). Um exemplo prático de aprendizado não-supervisionado é a identificação de espécies de insetos baseados num percentual de imagens usadas no treinamento de um sistema baseado em rede neural e um outro percentual de imagens que não foram usadas no treinamento para testar o aprendizado do sistema, possibilitando verificar o desempenho na identificação das espécies usadas no teste (Valan *et al.*, 2019).

Métodos de identificação automática de insetos

Com a finalidade de facilitar a correta identificação das espécies, métodos para identificação automatizada de organismos baseados na utilização de imagens têm sido desenvolvidos nas últimas décadas (Weeks *et al.*, 1997, 1999; Watson *et al.*, 2003; Macleod *et al.*, 2010; Wang *et al.*, 2012a, 2012b; Yang *et al.*, 2015; Wäldche & Mäder, 2018; Bambil *et al.*, 2020). Para os insetos, uma rápida identificação, tanto para espécies quanto para outras categorias taxonômicas, é importante em muitos contextos, visto que estes organismos representam a maior parte da diversidade biológica do planeta. Dessa forma, é necessário entender como está organizada a

composição e funcionamento dos ecossistemas, aperfeiçoando a capacidade de identificar efetivamente os insetos que os habitam (Valan *et al.*, 2019). Além da preocupação em facilitar a identificação precisa de determinados grupos para resolução de problemas taxonômicos, existe também a necessidade do desenvolvimento de métodos que facilitem a identificação de insetos de importância médica e econômica, objetivando o desenvolvimento de estratégias de controle de pragas, espécies invasoras e a vigilância entomológica de insetos transmissores de doenças (Cho *et al.*, 2007; Gurgel-Gonçalves *et al.*, 2017; Valan *et al.*, 2019).

Mesmo com os métodos tradicionais, a identificação de insetos é dificultada pela grande diversidade, com mais de um milhão de espécies descritas (Zhang, 2011), além de variações dentro das espécies relacionadas ao sexo, coloração, fases de desenvolvimento, dentre outras características que muitas vezes impossibilitam a identificação morfológica dos grupos (Valan *et al.*, 2019). Com a aquisição de treinamento, é possível aprender como identificar grupos de níveis taxonômicos mais altos, como ordens, porém essa tarefa se torna desafiadora quando diminuimos o nível hierárquico das categorias, no caso de famílias, gêneros e espécies. Dessa forma, quanto menor for o nível taxonômico, mais difícil a identificação precisa (Valan *et al.*, 2019).

Algumas contribuições clássicas na identificação automática de insetos incluem: 1) um sistema semiautomático desenvolvido e treinado com 175 imagens digitais, que permitiu discriminar cinco espécies fortemente relacionadas de himenópteros da família Ichneumonidae Latreille, 1802, utilizando como comparativo o padrão de asas alcançando uma identificação das espécies com até 94% de acurácia (Weeks *et al.*, 1997); 2) um sistema chamado DAISY (the Digital Automated Identification System), capaz de identificar espécies dos gêneros *Culicoides* Latreille, 1909 e *Forcipomyia* Meigen, 1818 (Diptera: Ceratopogonidae), buscando diminuir as limitações rotineiras enfrentadas pelos taxonomistas especializados na identificação de espécies. A acurácia alcançada foi de 86% para 49 espécies analisadas (Weeks *et al.*, 1999). 3) um sistema para identificar 237 espécies da ordem Lepidoptera Linnaeus, 1758 (especificamente macrolepidópteros), considerando iluminação inconsistente, variação na postura de repouso do inseto e inclusão de indivíduos com estruturas corpóreas danificadas, gerando uma grande variabilidade no conjunto de imagens. Os resultados na identificação das espécies variaram de 35% a 100% de identificações corretas. Os autores argumentaram que as imagens com baixa taxa de

identificações corretas foram as de indivíduos que perderam escamas ou que tiveram iluminação irregular, mostrando a importância na qualidade das imagens para identificações precisas (Watson, *et al.*, 2004). Além de aplicações na identificação a nível de espécie, outros sistemas também foram desenvolvidos para identificação em outras categorias taxonômicas, por exemplo, famílias da ordem Lepidoptera (Wang *et al.*, 2012a)

Com o avanço das técnicas de aprendizado de máquina, técnicas mais sofisticadas como redes neurais artificiais (ANNs – Artificial Neural Networks) e o algoritmo máquina de vetores de suporte (SVM – Support Vector Machine) chegaram a obter uma acurácia de até 93% na identificação de nove ordens e subordens de insetos intimamente relacionadas e quando o SVM foi aplicado a acurácia foi ainda maior, atingindo até 100% para uma das ordens avaliadas (Wang *et al.*, 2012b). O SVM também mostrou bom desempenho ao identificar uma amostra de 120 indivíduos de sete espécies da ordem Neuroptera Linnaeus, 1758. Para cinco repetições do experimento, a média de identificação de cada espécie variou de 90% a 98% de acurácia, sendo esta aumentada quando o conjunto de dados foi dividido em dois grupos, baseando-se em características específicas dos olhos compostos, alcançando 99% de acurácia (Yang *et al.*, 2015).

Outras abordagens são o uso de redes neurais convolucionais (CNNs – Convolutional Neural Networks) que podem extrair recursos relevantes automaticamente sem a intervenção humana, obtendo identificações precisas no nível de especialistas (Valan *et al.*, 2019). Além disso, essa técnica pode utilizar transferência de recursos, ou seja, uma rede neural convolucional pré-treinada com outro conjunto de dados que pode ser aplicada em nova tarefa de identificação de insetos. Um exemplo é a rede VGG16, algoritmo pré-treinado no conjunto de dados ImageNet. Essa rede foi utilizada para avaliar tarefas taxonômicas desafiadoras, como, 1) a identificação de insetos para grupos superiores quando eles pertenciam a subgrupos que não foram vistos anteriormente e 2) a identificação de espécies morfológicamente semelhantes que são difíceis de distinguir até mesmo para especialistas. Seu desempenho na identificação dos táxons variou de 92% a 98,6%, a depender do nível hierárquico dos grupos e da representatividade no número de imagens analisadas (Valan *et al.*, 2019).

Com as diversas propostas de desenvolvimento de métodos de identificação automática das espécies, MacLeod *et al.* (2010) sugerem que os taxonomistas devem

atuar com especialistas em reconhecimento de padrões, aprendizado de máquina e inteligência artificial, resultando em mais precisão e menos esforço na identificação dos organismos. Dessa forma, os taxonomistas estariam isentos dos desafios da identificação clássica, permitindo se dedicarem em questões e problemas mais difíceis de solucionar, como por exemplo, a descrição e revisão dos conceitos de espécies, o estudo dos papéis das espécies nos ecossistemas e suas relações filogenéticas (Macleod *et al.*, 2010).

Além da contribuição como ferramentas de taxonomia integrativa, o avanço das técnicas de aprendizado de máquina poderá contribuir também em outros diversos problemas da sociedade, como o monitoramento de insetos de importância agrícola e saúde pública (e.g. Liu *et al.*, 2016; Park *et al.*, 2020). Neste sentido, as redes neurais têm sido usadas também para a identificação de imagens de espécies de carrapatos, que são transmissores de bactérias que causam a febre maculosa e a doença de Lyme, por exemplo, alcançando uma acurácia de 87,8% nas identificações (Justen *et al.*, 2021). A aplicação de redes neurais convolucionais, como VGG, ResNet, Inception, MobileNet e DenseNet obtiveram acurácias superiores a 98% em tarefas de identificação de espécies de carrapatos [*Amblyomma americanum* (Linnaeus, 1758), *Dermacentor variabilis* (Say, 1821), e *Ixodes scapularis* Say, 1821], alcançando até 99,5% (Luo *et al.*, 2022). Outro grupo incluído nessa abordagem são os mosquitos, vetores de dengue, Chikungunya, Zika, filariose linfática e malária (Lorenz *et al.*, 2015; Motta *et al.*, 2019; 2020; Park *et al.*, 2020; Pataki *et al.*, 2021). Para mosquitos das espécies *Aedes aegypti* (Linnaeus, 1762), *Aedes albopictus* (Skuse, 1895) e *Culex quinquefasciatus* Say, 1823, as redes neurais LeNet, AlexNet e GoogleNet obtiveram acurácias de testes de 51,2%, 52,4% e 76,2%, respectivamente. Quando a tarefa envolvia apenas a identificação a nível de gênero, foram alcançadas acurácias de 100% e 90% para *Aedes* Meigen, 1818 e *Culex* Meigen, 1818, respectivamente (Motta *et al.*, 2019), mostrando que existia dificuldade para identificação de categorias taxonômicas mais específicas (Valan *et al.*, 2019). Em um estudo posterior a otimização de hiperparâmetros de modelos baseados em CNNs com arquiteturas complexas aumentou o desempenho de 76,2% para 93,5% de acurácia para as três espécies de mosquitos (Motta *et al.*, 2020). Em uma análise incluindo *Ae. albopictus*, *Ae. dorsalis* (Meigen, 1830), *Ae. koreicus* (Edwards, 1917), *Ae. vexans* (Meigen, 1830), *Anopheles* spp., *Cx. Inatomii* Kamimura and Wada, 1974, *Cx. pipiens* Linnaeus, 1758 e *Cx. tritaeniorhynchus* Giles, 1901 (sendo que *Ae.*

dorsalis, *Ae. koreikus* e *Cx. inatomii* foram considerados como uma única classe de menor potencial), o uso de três CNNs, VGG-16, ResNet-50 e SqueezeNet, alcançaram acurácias de 97%, 96% e 90%, respectivamente. Esse estudo utilizou mapas de recursos para ativar apenas as características que apresentavam influência positiva na predição das espécies, gerando um mapa de calor que mostra as principais regiões discriminativas nas imagens de entrada da rede (Park *et al.*, 2020).

Identificação automática de triatomíneos

Um projeto pioneiro de identificação automática de triatomíneos do México e Brasil foi baseado em uma rica base de dados obtida entre 2012-2017, com financiamento da Universidade do Kansas. O projeto desenvolveu um aparato para obtenção de fotografias dorsais de alta qualidade e padronizadas (Figura 6). Foi desenvolvido um sistema de identificação visual automática para 39 espécies de triatomíneos do Brasil e 12 espécies do México a partir do processamento de 1674 imagens, resultando em sucesso de identificação superior a 80%; quando se considerava as faunas locais nas análises os sucessos foram de aproximadamente 90% para as espécies brasileiras (Gurgel-Gonçalves *et al.*, 2017). Posteriormente, em um estudo utilizando os mesmos conjuntos de fotografias, e aplicando algoritmos baseados em redes neurais (deep learning), foi alcançado até 98,9% de sucesso na identificação das imagens para algumas espécies (Khalighifar *et al.*, 2019).

A construção de uma base de imagens bem organizada e com a supervisão de um taxonomista é fundamental para que algoritmos de aprendizado de máquina tenham um bom funcionamento, evitando falhas nas classificações de pré-treinamento (Gurgel-Gonçalves *et al.*, 2023) e com resultados que retratem a realidade das identificações. Um estudo usando o banco de imagens disponibilizado por Gurgel-Gonçalves *et al.* (2017), demonstrou um bom desempenho de algoritmos nas análises de identificações, porém ocorreram erros na rotulagem das classes analisadas e, conseqüentemente, erros na identificação dos triatomíneos (Parsons & Banitaan, 2021). Esse banco de imagens também foi usado na proposta de um modelo de aprendizagem profunda para diferenciar triatomíneos de não-triatomíneos (99,7% de acurácia), espécies de triatomíneos mexicanos (97,3%) e espécies de triatomíneos brasileiros (96,4%) (Abdelghani *et al.*, 2021). No entanto, os avanços nas aplicações de ferramentas promissoras, como as redes neurais e outros algoritmos de machine learning, podem mostrar um cenário positivo para o desenvolvimento de um sistema

acurado e eficaz para a identificação e vigilância dos vetores da doença de Chagas (Cochero *et al.*, 2022).

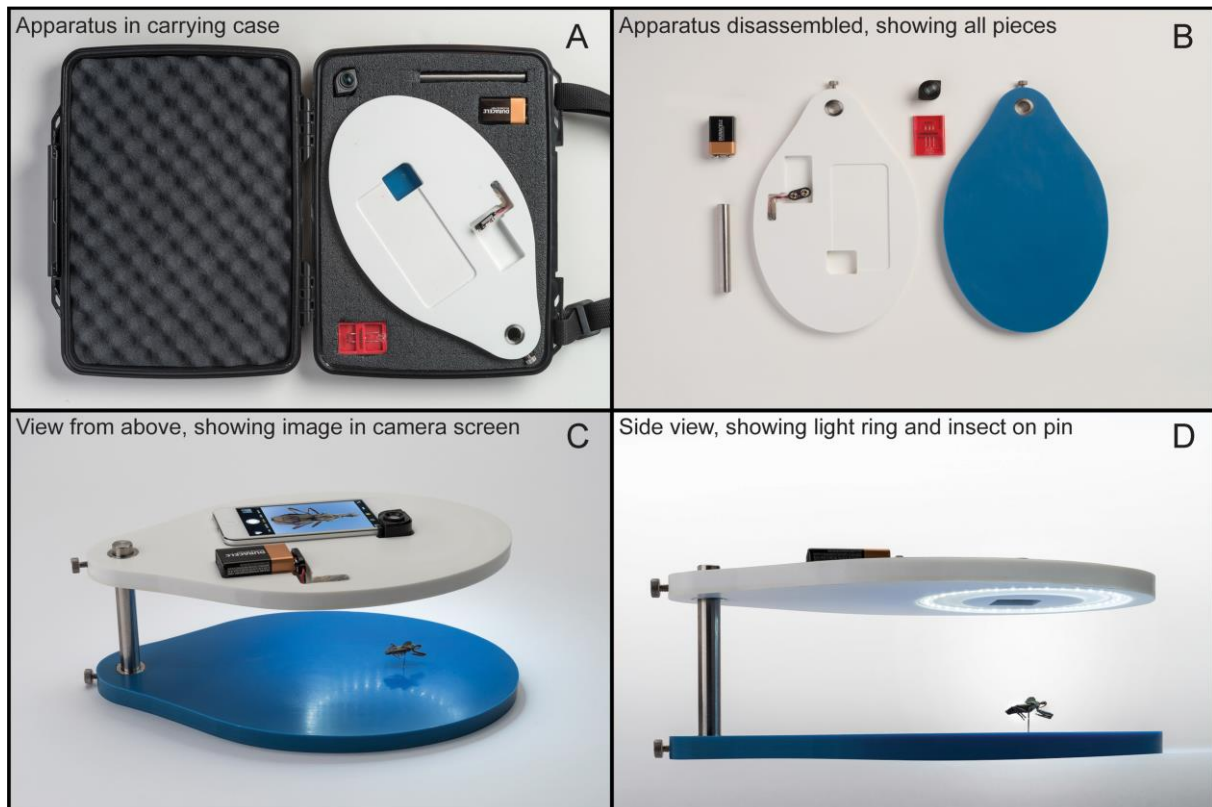


Figura 6. Equipamento fotográfico para obtenção das imagens de alta resolução de triatomíneos Fonte: Gurgel-Gonçalves *et al.*, 2017.

Sistemas com objetivo de identificar automaticamente triatomíneos a partir de fotografias seriam fundamentais para reforçar a vigilância dos vetores da doença de Chagas com participação comunitária, empoderando a população que convive com esses insetos, no contexto da ciência cidadã (Curtis-Robles *et al.*, 2015). Contudo, alguns pontos são importantes, como a validação taxonômica baseada em imagens, a geração de dados de treinamento maiores, o desenvolvimento de base de dados de referência que sejam públicos e com curadoria e que possam contribuir para a integração dos sistemas de identificação automática como uma ferramenta de taxonomia que seja utilizada pela população, fortalecendo a ciência cidadã (Høye *et al.*, 2021). Essas perspectivas poderão favorecer não apenas a vigilância e controle dos vetores da doença de Chagas (e.g. Delgado-Nogueira *et al.*, 2022), mas também o monitoramento de outros grupos de artrópodes com importância em saúde pública, com a ajuda comunitária (e.g. Hamer *et al.*, 2018; Sousa *et al.*, 2020; 2022; Gálvez *et*

al., 2022; Cull *et al.*, 2022; Ballman *et al.*, 2023). Embora muitas instituições de saúde pública registrem ocorrências de vetores, a maioria depende do envio de insetos para laboratórios de referência, e os dados resultantes não se tornam disponíveis em arquivos de dados abertos. Além disso, em muitos programas de vigilância de triatomíneos esses insetos não são corretamente identificados devido à rotatividade e treinamento insuficiente dos agentes de combate a endemias. Tem sido observado um declínio no número de taxonomistas especializados em vetores assim como uma mudança de foco dos estudos acadêmicos para áreas distintas de sua biologia, que podem comprometer a capacidade de resposta a crises de saúde pública (Carvalho *et al.*, 2005; Carvalho *et al.*, 2007; Casas *et al.*, 2016; Engel *et al.*, 2021). Dessa forma há a necessidade de desenvolver novas estratégias de identificação de triatomíneos para auxiliar a população a distinguir triatomíneos de outros insetos o que permitiria a efetiva vigilância entomológica com participação comunitária e a identificação de situações de potencial contato vetor-humano e risco de transmissão de *T. cruzi*.

Apesar de estudos prévios terem conseguido identificar espécies de triatomíneos automaticamente baseado em fotografias dorsais padronizadas de alta qualidade (Gurgel-Gonçalves *et al.* 2017; Khalighifar *et al.*, 2019), um passo importante para vigilância com participação comunitária seria diferenciar triatomíneos em diversas condições, já que a qualidade da imagem e as posições podem trazer e/ou esconder informações importantes para a identificação automática. Além disso, estudos anteriores mostraram que o sucesso de identificação depende dos algoritmos utilizados (Gurgel-Gonçalves *et al.* 2017, Khalighifar *et al.*, 2019), de maneira que são necessários estudos aplicando diferentes algoritmos (e.g. Valan *et al.*, 2019; Parsons & Banitaan, 2021) para avaliação da acurácia de identificação.

JUSTIFICATIVA

Esta tese apresenta inovações para identificar imagens de triatomíneos de maneira automática por meio de técnicas de aprendizagem de máquina. Um requisito fundamental para a vigilância da doença de Chagas é a identificação acurada dos triatomíneos. Atualmente, um inseto coletado passa por diferentes locais (casa do morador, posto de saúde, núcleo de entomologia) até chegar a um especialista que realiza a identificação, usando chaves dicotômicas impressas e eventualmente desatualizadas. Raramente as identificações são relatadas aos moradores e muitos vetores em potencial não são identificados corretamente. Dessa forma, há

necessidade de aprimorar as estratégias de identificação de triatomíneos para melhorar a vigilância dos vetores com participação comunitária, a qual é comprovadamente mais eficaz que a vigilância baseada nas atividades dos agentes de combate a endemias. O problema é que os sistemas de vigilância de vetores estão se tornando cada vez mais fracos na América Latina com a diminuição do número de especialistas em taxonomia de triatomíneos. Sistemas de vigilância fracos têm menor capacidade de estratificar o risco de transmissão vetorial e desenvolver estratégias de controle. A vigilância enfraquecida ainda produz menos registros de ocorrência de triatomíneos, o que leva a falsa impressão de que os vetores estão se tornando mais raros, subestimando o risco de transmissão vetorial da doença de Chagas.

Uma estratégia promissora para resolver esse problema é o desenvolvimento de novas ferramentas para fortalecer a vigilância de vetores, entre elas, o uso de sistemas baseados em aprendizado de máquina para a identificação automática de triatomíneos. Apesar de estudos prévios terem conseguido identificar automaticamente espécies de triatomíneos baseado em fotografias dorsais padronizadas de alta qualidade, uma inovação importante é identificar corretamente as imagens de triatomíneos (adultos ou ninfas) obtidas com câmeras digitais comuns em vários ângulos ou posições. Atualmente não existem sistemas de vigilância de triatomíneos com capacidade de identificação automática em países da América Latina. Funcionalidades incluem distinguir e identificar espécies de triatomíneos permitindo a efetiva vigilância dos vetores com participação comunitária, a identificação do potencial contato vetor-humano e a sinalização de risco de transmissão de *T. cruzi*. O desenvolvimento desse sistema permitirá gerar mais informações para a vigilância por meio da participação comunitária, contribuindo ainda para o desenvolvimento da ciência cidadã.

Para alcançar esse grande objetivo essa tese apresenta três capítulos com inovações para identificação automática dos triatomíneos. No primeiro, avaliamos o desempenho de cinco algoritmos de aprendizado de máquina para identificação de três gêneros de vetores da doença de Chagas com base em imagens em diferentes ângulos ou posições obtidas com uma câmera de celular comum. No segundo, avaliamos o desempenho da rede neural convolucional pré-treinada AlexNet na identificação de diversas espécies de triatomíneos a partir de imagens dorsais obtidas de celulares comuns e das imagens disponibilizadas por Gurgel-Gonçalves *et al.*

(2017). E finalmente no terceiro, avaliamos o desempenho da AlexNet na identificação de imagens de ninfas de triatomíneos até o nível de gênero.

OBJETIVOS

Geral

Avaliar o desempenho de técnicas de aprendizado de máquina para a identificação de imagens de triatomíneos adultos e ninfas e contribuir para a inovação da identificação automática de triatomíneos.

Específicos

- Obter imagens de espécies de triatomíneos e organizar um banco de imagens obtidas de câmeras de celulares com diferentes posições de fotografia.
- Avaliar o desempenho de cinco algoritmos de aprendizado de máquina para identificar imagens de triatomíneos de três gêneros em diferentes posições, obtidas com uma câmera de celular comum.
- Avaliar o desempenho de um algoritmo de aprendizagem profunda (AlexNet) para identificação de imagens dorsais de adultos de espécies de triatomíneos.
- Avaliar o desempenho do algoritmo AlexNet para identificação de imagens de ninfas de triatomíneos até o nível de gênero.

REFERÊNCIAS

- Abad-Franch, F., Diotaiuti, L., Gurgel-Gonçalves, R., & Gürtler, R. E. (2013). Certifying the interruption of Chagas disease transmission by native vectors: cui bono?. *Memórias do Instituto Oswaldo Cruz*, 108, 251-254.
- Abad-Franch, F. (2016). A simple, biologically sound, and potentially useful working classification of Chagas disease vectors. *Memórias do Instituto Oswaldo Cruz*, 111, 649-651.
- Abad-Franch, F., & Gurgel-Gonçalves, R. (2021). The ecology and natural history of wild Triatominae in the Americas. In Guarneri, A & Lorenzo, M (Ed.). *Triatominae - The Biology of Chagas Disease Vectors*. (Vol. 5, Chap. 16, pp. 387-445). Cham: Springer International Publishing.
- Abdelghani, B. A., Banitaan, S., Maleki, M., & Mazen, A. (2021). Kissing bugs identification using convolutional neural network. *IEEE Access*, 9, 140539-140548.
- Alevi, K. C. C., de Oliveira, J., da Silva Rocha, D., & Galvão, C. (2021). Trends in taxonomy of Chagas disease vectors (Hemiptera, Reduviidae, Triatominae): From Linnaean to integrative taxonomy. *Pathogens*, 10(12), 1627.
- Ballman, E. S., Leahy, J. E., Sponarski, C. C., Galli, M. G., & Gardner, A. M. (2023). A citizen science approach to investigate the distribution, abundance, and pathogen infection of vector ticks through active surveillance. *Ticks and Tick-borne Diseases*, 14(3), 102144.
- Bambil, D., Pistori, H., Bao, F., Weber, V., Alves, F. M., Gonçalves, E. G., Figueiredo, L. F. A., Abreu, U. G. P., Arruda, R., & Bortolotto, I. M. (2020). Plant species identification using color learning resources, shape, texture, through machine learning and artificial neural networks. *Environment Systems and Decisions*, 40(4), 480-484.
- Bartlett, C. R., Deitz, L. L., Dmitriev, D. A., Sanborn, A. F., Soulier-Perkins, A., & Wallace, M. S. (2018). The diversity of the true hoppers (Hemiptera: Auchenorrhyncha). *Insect biodiversity: science and society*, 2, 501-590.
- Brito, R. N., Diotaiuti, L., Celencina Fagundes Gomes, A., de Cássia Moreira de Souza, R., & Abad-Franch, F. (2017). *Triatoma costalimai* (Hemiptera: Reduviidae) in and

- around houses of Tocantins State, Brazil, 2005–2014. *Journal of Medical Entomology*, 54(6), 1771-1774.
- Brito, R. N., Gorla, D. E., Diotaiuti, L., Gomes, A. C., Souza, R. C., & Abad-Franch, F. (2017). Drivers of house invasion by sylvatic Chagas disease vectors in the Amazon-Cerrado transition: A multi-year, state-wide assessment of municipality-aggregated surveillance data. *PLoS neglected tropical diseases*, 11(11), e0006035.
- Casas, J., Lazzari, C., Insausti, T., Launois, P., & Fouque, F. (2016). Mapping of courses on vector biology and vector-borne diseases systems: time for a worldwide effort. *Memórias do Instituto Oswaldo Cruz*, 111, 717-719.
- Carcavallo, R. U., Galindez, I., José, J., & Herman, L. (1997). *Atlas dos vetores da doença de Chagas nas Américas*. Rio de Janeiro: Fiocruz Editorial.
- Carvalho, M. R., Bockmann, F. A., Amorim, D. S., de Vivo, M., de Toledo-Piza, M., Menezes, N. A., Figueredo, J. L., Compagno, L. J. V., Schelly, R. C., Britz, R., Lundberg, J. G., Vari, R. P., & Nelson, G. (2005). Revisiting the taxonomic impediment. *Science*, 307(5708), 353-353.
- Carvalho, M. R., Bockmann, F. A., Amorim, D. S., Brandão, C. R. F., de Vivo, M., de Figueiredo, J. L., Britski, H. A., Pinna, M. C. C., Menezes, N. A., Marques, F. P. R., Papavero, N., Canello, E. M., Crisci, J. V., McEachran, J. D., Schelly, R. C., Lundberg, J. G., Gill, A. C., Britz, R., Wheeler, Q. D., Stiassny, M. L. J., Parenti, L. R., Page, L. M., Wheeler, W. C., Faivovich, & Nelson, G. J. (2007). Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, 34, 140-143.
- Cavalier-Smith, T. (2016). Higher classification and phylogeny of Euglenozoa. *European Journal of Protistology*, 56, 250-276.
- Cho, J., Choi, J., Qiao, M., Ji, C. W., Kim, H. Y., Uhm, K. B., & Chon, T. S. (2007). Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis. *Red*, 346(246), 244.
- Cochero, J., Pattori, L., Balsalobre, A., Ceccarelli, S., & Marti, G. (2022). A convolutional neural network to recognize Chagas disease vectors using mobile phone images. *Ecological Informatics*, 68, 101587.

- Cohen, A. C. (1990). Feeding adaptations of some predaceous Hemiptera. *Annals of the Entomological Society of America*, 83(6), 1215-1223.
- Cruz, D. D., Arellano, E., Denis Ávila, D., & Ibarra-Cerdeña, C. N. (2020). Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: a case for the cryptic dimidiata complex. *Parasites & vectors*, 13, 1-12.
- Cull, B. (2022). Monitoring trends in distribution and seasonality of medically important ticks in North America using online crowdsourced records from iNaturalist. *Insects*, 13(5), 404.
- Curtis-Robles, R., Wozniak, E. J., Auckland, L. D., Hamer, G. L., & Hamer, S. A. (2015). Combining public health education and disease ecology research: using citizen science to assess Chagas disease entomological risk in Texas. *PLoS neglected tropical diseases*, 9(12), e0004235.
- Dayrat, B. (2005). Towards integrative taxonomy. *Biological journal of the Linnean society*, 85(3), 407-417.
- Delgado-Noguera, L. A., Hernández-Pereira, C. E., Ramírez, J. D., Hernández, C., Velasquez-Ortiz, N., Clavijo, J., Ayala, J. M., Forero-Peña, D., Marquez, M., Suarez, M. J., Traviezo-Valles, L., Escalona, M. A., Perez-Garcia, L., Carpio, I. M., Sordillo, E. M., Grillet, M. E., Llewellyn, M. S., Gabaldón, J. C., & Mondolfi, A. E. P. (2022). Tele-entomology and tele-parasitology: A citizen science-based approach for surveillance and control of Chagas disease in Venezuela. *Parasite Epidemiology and Control*, 19, e00273.
- Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920-1930.
- Dias, J. C. P. (2007). Southern Cone Initiative for the elimination of domestic populations of *Triatoma infestans* and the interruption of transfusion Chagas disease: historical aspects, present situation, and perspectives. *Memórias do Instituto Oswaldo Cruz*, 102, 11-18.
- Dujardin, J. P., Panzera, P., & Schofield, C. J. (1999). Triatominae as a model of morphological plasticity under ecological pressure. *Memórias do Instituto Oswaldo Cruz*, 94, 223-228.
- Engel, M. S., Ceríaco, L. M., Daniel, G. M., Dellapé, P. M., Löbl, I., Marinov, M., ... & Zacharie, C. K. (2021). The taxonomic impediment: a shortage of taxonomists, not

- the lack of technical approaches. *Zoological Journal of the Linnean Society*, 193(2), 381-387.
- Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, 37(2), 505-515.
- Gallo, D., Nakano, O., Silveira Neto, S., Carvalho, R., Baptista, G., Berti Filho, E., Parra, J. R. P., Zucchi, R. A., Zucchi, R. A., Alves, S. B., Vendramim, J. D., Marchini, L. C., Lopes, J. R. S., & Omoto, C. (2002). *Entomologia agrícola*. Piracicaba: FEALQ.
- Galvão, C. (2014). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia. Curitiba: Sociedade Brasileira de Zoologia.
- Galvão, C., & Dale, C. (2014). Chaves de identificação para adultos. In Galvão, C. (Org.). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia.
- Galvão, C. (2003). A sistemática dos triatomíneos (Hemiptera, Reduviidae), de De Geer ao DNA. *Entomología y Vectores*, 10(4), 511-30.
- Goodchild, A. J. P. (1966). Evolution of the alimentary canal in the Hemiptera. *Biological Reviews*, 41(1), 97-139.
- Gálvez, R., de Felipe, M. L., & Yebes, F. (2022). Citizen science set in motion: DIY light traps for phlebotomine sand flies. *Preventive Veterinary Medicine*, 200, 105589.
- Gullan, P. J., & Cook, L. G. (2007). Phylogeny and higher classification of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Zootaxa*, 1668(1), 413-425.
- Gurgel-Gonçalves, R., Komp, E., Campbell, L. P., Khalighifar, A., Mellenbruch, J., Mendonça, V. J., Owens, H. L., Felix, K. L. C., Peterson, A. T., & Ramsey, J. M. (2017). Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the Virtual Vector Lab. *PeerJ*, 5, e3040.
- Gurgel-Gonçalves, R., Abad-Franch, F., de Almeida, M. R., Obara, M. T., de Souza, R. D. C. M., Batista, J. A. D. S., & Rocha, D. D. A. (2021). TriatoDex, an electronic identification key to the Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease: Development, description, and performance. *Plos one*, 16(4), e0248628.

- Gurgel-Gonçalves, R., Miranda, V. L., Khalighifar, A., & Peterson, A. T. (2023). Shooting in the dark: automatic identification of disease vectors without taxonomic expert supervision. *Ecological Informatics*, 75, 102029.
- Hamer, S. A., Curtis-Robles, R., & Hamer, G. L. (2018). Contributions of citizen scientists to arthropod vector data in the age of digital epidemiology. *Current opinion in insect science*, 28, 98-104.
- Hashimoto, K., & Schofield, C. J. (2012). Elimination of *Rhodnius prolixus* in central America. *Parasites & vectors*, 5, 1-10.
- Høye, T. T., Ärje, J., Bjerger, K., Hansen, O. L., Iosifidis, A., Leese, F., Mann, H. M. R., Meissner, K., Melvad, C., & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, 118(2), e2002545117.
- Jeannel, R. (1919). Insectes Hémiptères III. Hemicocephalidae et Reduviidae. In: *Voyage de Ch. Allaucaud et R. Jeannel en Afrique Orientale (1911-1912)*. Paris: Résultats Scientifiques.
- Jurberg, J., & Galvão, C. (1997). Hermanlenticia n. gen. da tribo Triatomini, com um rol de espécies de Triatominae (Hemiptera, Reduviidae). *Memórias do Instituto Oswaldo Cruz*, 92, 181-185.
- Jurberg, J., Galvão, C., François, N., Carcavallo, R. U., Rocha, D. S., & Lent, H. (2005). *Uma iconografia dos Triatomíneos*. Rio de Janeiro: Co-edição Hill Edirions e Editora Gama Filho.
- Justen, L., Carlsmith, D., Paskewitz, S. M., Bartholomay, L. C., & Bron, G. M. (2021). Identification of public submitted tick images: A neural network approach. *Plos one*, 16(12), e0260622.
- Khalighifar, A., Komp, E., Ramsey, J. M., Gurgel-Gonçalves, R., & Peterson, A. T. (2019). Deep learning algorithms improve automated identification of Chagas disease vectors. *Journal of medical entomology*, 56(5), 1404-1410.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1), 89-109.
- Lazzari, C. R. (2014). Biologia e Comportamento. In Galvão, C. (Org.). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia.

- Lehmann, P., Ordoñez, R., Ojeda-Baranda, R., Lira, J., Hidalgo-Sosa, L., Monroy, C., & Ramsey, J. M. (2005). Morphometric analysis of *Triatoma dimidiata* populations (Reduviidae: Triatominae) from Mexico and northern Guatemala. *Memórias do Instituto Oswaldo Cruz*, 100, 477-486.
- Lent, H., & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bulletin of the American museum of Natural History*, 163(3), 123-520.
- Liu, Z., Gao, J., Yang, G., Zhang, H., & He, Y. (2016). Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific reports*, 6(1), 20410.
- Lorenz, C., Ferraudo, A. S., & Suesdek, L. (2015). Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Tropica*, 152, 165-169.
- Lukashevich, E. D., & Mostovski, M. B. (2003). Hematophagous insects in the fossil record. *Paleontological Journal*, 37(2), 153-161.
- Luo, C. Y., Pearson, P., Xu, G., & Rich, S. M. (2022). A Computer Vision-Based Approach for Tick Identification Using Deep Learning Models. *Insects*, 13(2), 116.
- MacLeod, N., Benfield, M., & Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312), 154-155.
- Magoulas, G. D., & Prentza, A. (1999). Machine learning in medical applications. In: G. Paliouras., V. Karkaletsis., & C. D. Spyropoulos (Eds.). *Machine Learning and Its Applications*. Berlin: Springer Berlin Heidelberg.
- Maldonado-Capriles, J. (1990). Systematic catalogue of the Reduviidae of the world (Insecta: Heteroptera). *Caribbean Journal of Science*, 1-694.
- Mendonça, V. J., da Silva, M. T., de Araújo, R. F., Júnior, J. M., Júnior, M. B., Almeida, C. E., Costa, J., Graminha, M. A. S., Cicarelli, R. M. B., & Rosa, J. A. (2009). Phylogeny of *Triatoma sherlocki* (Hemiptera: Reduviidae: Triatominae) inferred from two mitochondrial genes suggests its location within the *Triatoma brasiliensis* complex. *The American journal of tropical medicine and hygiene*, 81(5), 858-864.
- Monroy, C., Bustamante, D. M., Rodas, A., Rosales, R., Mejía, M., & Tabaru, Y. (2003). Geographic distribution and morphometric differentiation of *Triatoma nitida* Usinger

- 1939 (Hemiptera: Reduviidae: Triatominae) in Guatemala. *Memórias do Instituto Oswaldo Cruz*, 98, 37-43.
- Monteiro, F. A., Weirauch, C., Felix, M., Lazoski, C., & Abad-Franch, F. (2018). Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. *Advances in parasitology*, 99, 265-344.
- Motta, D., Santos, A. Á. B., Winkler, I., Machado, B. A. S., Pereira, D. A. D. I., Cavalcanti, A. M., Fonseca, E. O. L., Kirchner, F., & Badaró, R. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. *PloS one*, 14(1), e0210829.
- Motta, D., Santos, A. Á. B., Machado, B. A. S., Ribeiro-Filho, O. G. V., Camargo, L. O. A., Valdenegro-Toro, M. A., Kirchner, F., & Badaró, R. (2020). Optimization of convolutional neural network hyperparameters for automatic classification of adult mosquitoes. *Plos one*, 15(7), e0234959.
- Noireau, F., & Dujardin, J. P. (2010). Biology of Triatominae. In: Jansen, A. M., & Roque, A. L. R. (Orgs.). *American Trypanosomiasis Chagas Disease: One Hundred Years of Research*. Amsterdam: Elsevier.
- Oliveira, L. M., Brito, R. N., Guimarães, P. A. S., Santos, R. V. M. A., Diotaiuti, L. G., Souza, R. C. M., & Ruiz, J. C. (2017). TriatoKey: a web and mobile tool for biodiversity identification of Brazilian triatomine species. *Database*, 2017, bax033.
- Panizzi, A. R., & Grazia, J. (2015). Introduction to True Bugs (Heteroptera) of the Neotropics. In: Panizzi, A., & Grazia, J. (Eds.). *True Bugs (Heteroptera) of the Neotropics*. Springer: Dordrecht.
- Park, J., Kim, D. I., Choi, B., Kang, W., & Kwon, H. W. (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Scientific reports*, 10(1), 1012.
- Parsons, Z., & Banitaan, S. (2021). Automatic identification of Chagas disease vectors using data mining and deep learning techniques. *Ecological Informatics*, 62, 101270.
- Pataki, B. A., Garriga, J., Eritja, R., Palmer, J. R., Bartumeus, F., & Csabai, I. (2021). Deep learning identification for citizen science surveillance of tiger mosquitoes. *Scientific reports*, 11(1), 4718.

- Pinto, C. (1926). Classificação dos Triatomídeos (Hemiptera-Heteroptera hematofagos). *Scientia Medica*, 9, 485-490.
- Ribeiro Jr, G., Gurgel-Goncalves, R., Reis, R. B., Santos, C. G. S. D., Amorim, A., Andrade, S. G., & Reis, M. G. (2015). Frequent house invasion of *Trypanosoma cruzi*-infected triatomines in a suburban area of Brazil. *PLoS Neglected Tropical Diseases*, 9(4), e0003678.
- Rocha, D. D. A., Almeida, M. R., Batista, J., & Andrade, A. J. (2019). LutzDex™—A digital key for Brazilian sand flies (Diptera, Phlebotominae) within an Android App. *Zootaxa*, 4688(3), 382-388.
- Rojas de Arias, A., Monroy, C., Guhl, F., Sosa-Estani, S., Santos, W. S., & Abad-Franch, F. (2022). Chagas disease control-surveillance in the Americas: the multinational initiatives and the practical impossibility of interrupting vector-borne *Trypanosoma cruzi* transmission. *Memórias do Instituto Oswaldo Cruz*, 117, e210130.
- Roque, A. L. R., & Jansen, A. M. (2014). Reservatórios do *Trypanosoma cruzi* e sua relação com os vetores. In Galvão, C. (Org.). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia.
- Schmunis, G. A. (2000). A Tripanossomíase Americana e seu impacto na saúde pública das Américas. *Trypanosoma cruzi e Doença de Chagas*, 2, 1-15.
- Silveira, A. C., & Martins, E. (2014). Histórico do controle da transmissão vetorial e situação epidemiológica atual. In Galvão, C. (Org.). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia.
- Sherlock, I. A., Carcavallo, R. U., & Girón, I. G. (1997). Lista de infecções naturais e experimentais por flagelados em diversas espécies de triatomíneos. In: Carcavallo, R. U., Girón, I. G., Jurberg, J., & Lent, H. (Orgs.). *Atlas dos vetores da doença de Chagas nas Américas*. Fiocruz: Rio de Janeiro.
- Schuh, R. T., & Slater, J. A. (1995). *True bugs of the world (Hemiptera: Heteroptera): classification and natural history*. New York: Cornell University press.
- Sousa, L. B., Fricker, S. R., Doherty, S. S., Webb, C. E., Baldock, K. L., & Williams, C. R. (2020). Citizen science and smartphone e-entomology enables low-cost

- upscaling of mosquito surveillance. *Science of the Total Environment*, 704, 135349.
- Sousa, L. B., Craig, A., Chitkara, U., Fricker, S., Webb, C., Williams, C., & Baldock, K. (2022). Methodological diversity in citizen science mosquito surveillance: a scoping review. *Citizen Science: Theory and Practice*, 7(1).
- Souza, R. D. C. M. D., Brito, R. N., Barbosa, A. B., & Diotaiuti, L. (2014). Species of the subfamily Triatominae Jeannel, 1919 (Hemiptera: Reduviidae) present in the collection of Chagas disease vectors (FIOCRUZ-COLVEC), State of Minas Gerais. *Revista da Sociedade Brasileira de Medicina Tropical*, 47, 728-738.
- Ung, V., Dubus, G., Zaragüeta-Bagils, R., & Vignes-Lebbe, R. (2010). Xper2: introducing e-taxonomy. *Bioinformatics*, 26(5), 703-704.
- Usinger, R. L. (1943). A revised classification of the Reduvidae with a new subfamily from South America (Hemiptera). *Annals of the Entomological Society of America*, 36(4), 602-618.
- Valan, M., Makonyi, K., Maki, A., Vondráček, D., & Ronquist, F. (2019). Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology*, 68(6), 876-895.
- Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11), 2216-2225.
- Waleckx, E., Gourbière, S., & Dumonteil, E. (2015). Intrusive versus domiciliated triatomines and the challenge of adapting vector control practices against Chagas disease. *Memórias do Instituto Oswaldo Cruz*, 110, 324-338.
- Walter, D. E., & Winterton, S. (2007). Keys and the crisis in taxonomy: extinction or reinvention?. *Annual Review of Entomology*, 52, 193-208.
- Wang, J., Ji, L., Liang, A., & Yuan, D. (2012). The identification of butterfly families using content-based image retrieval. *Biosystems engineering*, 111(1), 24-32.
- Wang, J., Lin, C., Ji, L., & Liang, A. (2012). A new automatic identification system of insect images at the order level. *Knowledge-Based Systems*, 33, 102-110.

- Watson, A. T., O'Neill, M. A., & Kitching, I. J. (2004). Automated identification of live moths (Macrolepidoptera) using digital automated identification System (DAISY). *Systematics and Biodiversity*, 1(3), 287-300.
- Weeks, P. J. D., Gauld, I. D., Gaston, K. J., & O'Neill, M. A. (1997). Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research*, 87(2), 203-211.
- Weeks, P. J. D., O'Neill, M. A., Gaston, K. J., & Gauld, I. D. (1999). Automating insect identification: exploring the limitations of a prototype system. *Journal of Applied Entomology*, 123(1), 1-8.
- Weirauch, C. (2008). Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on morphological characters. *Systematic Entomology*, 33(2), 229-274.
- Weirauch, C., Bérenger, J. M., Berniker, L., Forero, D., Forthman, M., Frankenberg, S., Freedman, A., Gordon, E., Hoey-Chamberlain, R., Hwang, W. S., Marshall, S. A., Michael, A., Paiero, S. M., Udah, O., Watson, C., Yeo, M., Zhang, G., & Zhang, J. (2014). An illustrated identification key to assassin bug subfamilies and tribes (Hemiptera: Reduviidae). *Canadian Journal of Arthropod Identification*, 26(2), 1-115.
- Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine learning in medical imaging. *IEEE signal processing magazine*, 27(4), 25-38.
- WHO - World Health Organization. *Chagas disease (American trypanosomiasis)*. (2019). Disponível em <http://www.who.int/mediacentre/factsheets/fs340/en/>. Acesso em: 11 ago. 2020.
- Yang, H. P., Ma, C. S., Wen, H., Zhan, Q. B., & Wang, X. L. (2015). A tool for developing an automatic insect identification system based on wing outlines. *Scientific reports*, 5(1), 12786.
- Zhao, Y., Fan, M., Li, H., & Cai, W. (2023). Review of Kissing Bugs (Hemiptera: Reduviidae: Triatominae) from China with Descriptions of Two New Species. *Insects*, 14(5), 450.
- Zhang, Z. Q. (2011). *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*. New Zealand: Magnolia press.

CAPÍTULO I

Cellphone picture-based, genus-level automated identification of Chagas disease vectors: effects of picture orientation on the performance of five machine-learning algorithms

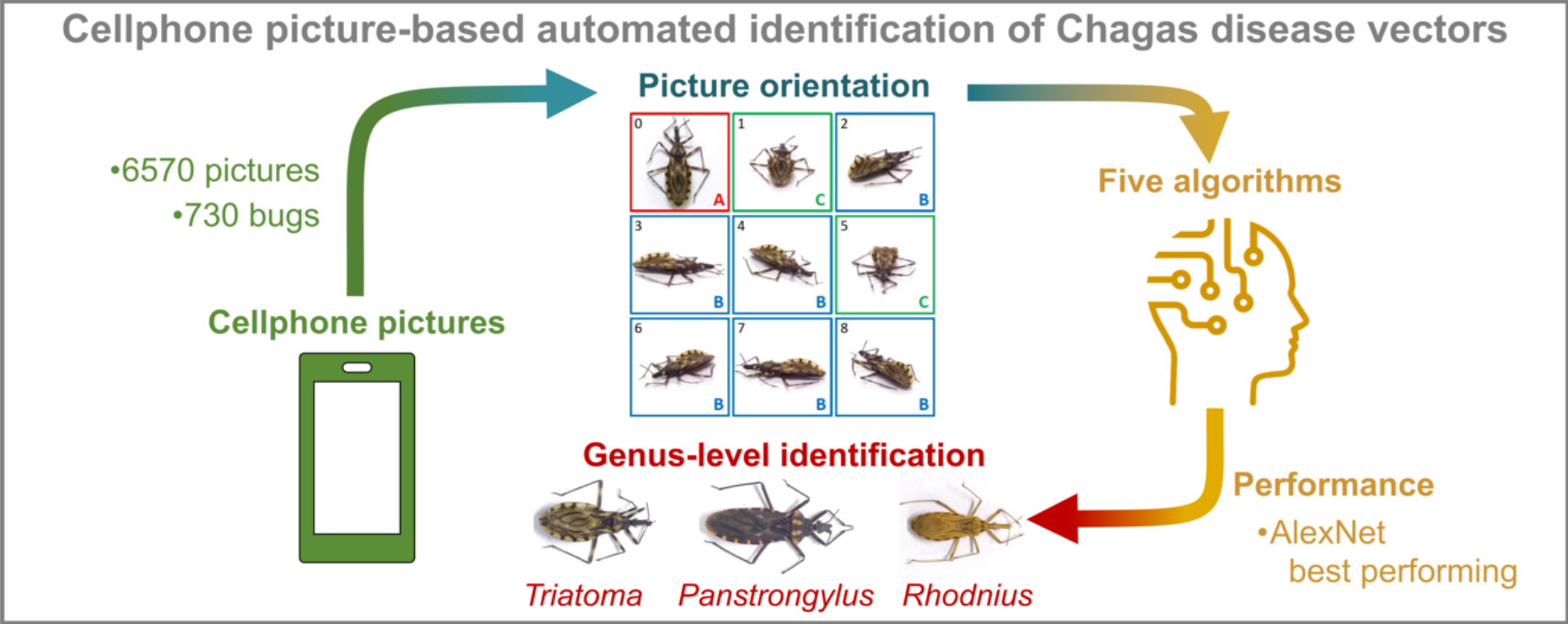
Manuscrito submetido à revista Ecological Informatics

(Apresentado segundo as normas da revista)

Highlights

- Chagas disease control requires vector surveillance, yet current systems are weak
- Automated vector identification may help strengthen surveillance
- To be useful, algorithms must work well with low-quality cellphone pictures
- Neural network accurately identified vector genera using low-resolution pictures
- Picture orientation (angle/position) barely affected genus-level identification

Graphical Abstract



Cellphone picture-based, genus-level automated identification of Chagas disease vectors: effects of picture orientation on the performance of five machine-learning algorithms

Vinícius Lima de Miranda^{a,b}, Ewerton Pacheco de Souza^b, Deborah Bambil^c, Ali Khalighifar^d, Andrew Townsend Peterson^e, Francisco Assis de Oliveira Nascimento^f, Rodrigo Gurgel-Gonçalves^{b*†}, Fernando Abad-Franch^{b,g†}

^a Departamento de Zoologia, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Distrito Federal, Brazil

^b Faculdade de Medicina, Universidade de Brasília, Brasília, Distrito Federal, Brazil

^c Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Distrito Federal, Brazil

^d Fish, Wildlife, and Conservation Biology Department, Colorado State University, Fort Collins, Colorado, USA

^e Department of Ecology and Evolutionary Biology and Biodiversity Institute, University of Kansas, Lawrence, Kansas, USA

^f Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil

^g Instituto Leônidas e Maria Deane – Fiocruz Amazônia, Manaus, Amazonas, Brazil

*Corresponding author: gurgelrg@hotmail.com

†Equal contribution

Abstract

Chagas disease is a major public-health concern across Latin America. It is caused by *Trypanosoma cruzi*, a parasite transmitted by blood-sucking triatomine bugs; the main vectors belong in three genera – *Triatoma*, *Panstrongylus*, and *Rhodnius*. Automated identification of triatomine bugs is a potential means to strengthen Chagas disease vector surveillance. To be broadly useful, however, automated systems must rely on algorithms capable of correctly identifying bugs pictured with ordinary digital cameras at varying angles or positions. Here, we assess the performance of five machine-learning algorithms at identifying the three main Chagas disease vector genera based on bugs pictured at different angles/positions with an ordinary cellphone camera. Each of 730 bugs (13 species) was pictured (72 dpi) at nine angles representing three positions: dorsal-flat, dorsal-oblique, and front/back-oblique. We randomly split the 6570-picture database into training (80%) and testing sets (20%), and then trained and tested a pre-trained convolutional neural network (AlexNet, AN); three boosting-based classifiers (AdaBoost, AB; Gradient Boosting, GB; and Histogram-based Gradient Boosting, HB); and a linear discriminant model (LD). To gauge performance consistency, each algorithm was trained-tested in 10 (pseudo)-replicate runs. We assessed identification accuracy and specificity with logit-binomial generalized linear mixed models (GLMMs) including fixed (algorithm; bug genus; picture angle or bug position) and random effects (picture; specimen; (pseudo)-replicate run). Models were fit in a Bayesian framework and evaluated using information-theoretic criteria. Differences in performance across algorithms were mainly driven by AN's essentially perfect accuracy and specificity, irrespective of picture angle or bug position. HB predicted accuracies ranged from ~0.987 (*Panstrongylus*, dorsal-oblique) to >0.999 (*Triatoma*, dorsal-flat). AB accuracy was poor for *Rhodnius* (~0.224–0.282) and *Panstrongylus* (~0.664–0.729), but fair for *Triatoma* (~0.988–0.991). For *Panstrongylus*, LD and GB had predicted accuracies in the ~0.970–0.984 range. AB

misclassified ~57% of *Rhodnius* and *Panstrongylus* as *Triatoma*, whereas specificity estimates ranged from ~0.92 to ~1.0 for the remaining algorithm-genus combinations. Dorsal-flat pictures appeared to slightly improve algorithm performance in some cases, but angle/position effects were overall weak-to-negligible. We conclude that, when high-performance machine-learning algorithms such as AN are used, the angles or positions at which bugs are pictured seem unlikely to hinder cellphone picture-based automated identification of Chagas disease vectors at the genus level. Future research should focus on combining mixed-quality pictures and pre-trained neural networks to (i) identify triatomine-bug adults to the species level and (ii) distinguish triatomine-bug nymphs (i.e., immature stages) from the nymphs of other, non-blood-feeding bugs.

Keywords: Automated identification; Machine learning; Accuracy; Specificity; Low-resolution pictures; Triatominae

Abbreviations: DALY, Disability-Adjusted Life Year; AN, AlexNet; AB, AdaBoost Multi-Class Adaptive Boosting classifier; GB, Gradient Boosting classifier; HB, Histogram-based Gradient Boosting classifier; LD, Linear Discriminant model; GLMM, Generalized Linear Mixed Model; AICc, Akaike's Information Criterion corrected for finite samples

Funding: This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), and the US National Science Foundation (NSF, United States of America; award number OIA-1920946).

1. Introduction

Pathogens transmitted by insects and other arthropods threaten public health globally; in 2019, an estimate ~700,000 people died, and ~53.2 million disability-adjusted life years (DALYs) were lost, due to diseases caused by insect-borne viruses (dengue, yellow fever, and Zika), bacteria (trachoma), or parasites (onchocerciasis, lymphatic filariasis, leishmaniasis, African trypanosomiasis, and Chagas disease) (GBD, 2019). In Latin America, the highest-burden vector-borne disease is Chagas disease, with recent estimates suggesting that its causative agent, *Trypanosoma cruzi*, infects ~5.5 to 7.3 million people and leads to annual losses of ~169,000 to 490,000 DALYs (GBD, 2019; Rojas de Arias et al., 2022).

Trypanosoma cruzi is a blood/tissue protozoan parasite that infects a wide range of mammalian hosts, including humans, and can cause severe heart, digestive, and, less frequently, neurological lesions (Prata, 2001; Jansen and Roque, 2010; Perez-Molina and Molina, 2018). *Trypanosoma cruzi* is endemic to the Americas, where it is primarily transmitted by a diverse group of blood-sucking true bugs known as triatomines (Lent and Wygodzinsky, 1979; Monteiro et al., 2018).

A critical requirement for the effective control-surveillance of vector-borne diseases, including Chagas disease, is accurate vector identification (Monteiro et al., 2018). However, current vector-surveillance systems are generally weak (and as a rule becoming even weaker) in Latin America and elsewhere, with expertise in disease-vector taxonomy overall declining (e.g., Yadon et al., 2006; Casas et al., 2016; Almeida et al., 2017; Gurgel-Gonçalves, 2022). Weak surveillance systems limit our ability to assess and stratify disease-transmission risk, which in turns hampers the planning and deployment of adequate public-health responses (Hashimoto et al., 2015; Yoshioka et al., 2017; Gürtler and Cecere, 2021). Weaker surveillance, moreover, yields progressively lower-quality, and sparser, vector-occurrence records (Abad-Franch et al., 2013, 2014). This not only hinders research aimed at

understanding how vector distribution patterns change over time or across space; worse still, it often creates the false impression that the vectors themselves, and not just the records of their occurrence, are becoming rarer – a ‘low-risk delusion’ that may self-reinforce *via* fading priority, funding cuts, even sparser records, and so forth in a perverse, positive-feedback loop (Schofield et al., 2006; Abad-Franch et al., 2013, 2014).

In this context of ever-weakening surveillance systems, researchers and public-health officials have insisted on the need to develop and test new tools and strategies to strengthen Chagas disease vector surveillance (e.g., WHO, 2002, 2017; Schofield et al., 2006; Abad-Franch et al., 2011, 2014; Dias et al., 2016; Gürtler and Cecere, 2021; Gurgel-Gonçalves, 2022). Efforts in that direction have addressed several sides of surveillance, including, e.g., improving vector-detection probabilities (Abad-Franch et al., 2011; Gürtler and Cecere, 2021); promoting instruction in medical entomology (Casas et al., 2016; Gurgel-Gonçalves, 2022); devising entomological-risk scoring/stratification systems that do not rely entirely on actual vector-occurrence records (e.g., Ribeiro-Jr et al., 2021); or developing electronic, easy-to-use vector identification keys (Oliveira et al., 2017; Gurgel-Gonçalves et al., 2021). An approach that has gained traction in recent years is using machine learning-based systems for the automated identification of triatomine bugs (Gurgel-Gonçalves et al., 2017; Khalighifar et al., 2019; Abdelghani et al., 2021; Cruz et al., 2020, 2021; Parsons and Banitaan 2021; Cocherro et al., 2022). In these systems, machine-learning algorithms are trained, then tested, on collections of digital pictures featuring bugs of known taxonomic identity. This ‘ground-truth’ picture set is usually composed of relatively high-quality, standardized digital pictures displaying the full dorsal habitus of the bugs (e.g., Gurgel-Gonçalves et al., 2017; Khalighifar et al., 2019; Abdelghani et al., 2021; Cruz et al., 2021). To be broadly useful, however, automated systems must rely on algorithms capable of correctly identifying bugs pictured with ordinary digital cameras at varying angles or positions – the kind of information that

surveillance staff or citizens in general can be foreseen to generate and feed into real-world, ‘extended’ vector-surveillance systems (Cochero et al., 2022; see also, e.g., Motta et al., 2019; Park et al., 2020; Terry et al., 2020; Justen et al., 2021; Pataki et al., 2021).

Here, we assess the performance of five machine-learning algorithms at identifying Chagas disease vectors based on bugs pictured at different angles or positions with an ordinary cellphone camera. Our assessment of algorithm performance builds upon statistical models that formally account for the non-independence of results arising from the same individual picture, from pictures featuring the same individual specimen, and from (pseudo-)replicate training-testing runs in which all algorithms used the same training and testing picture sets (see below); thus, model-based algorithm performance estimates and their associated measures of uncertainty both take data dependencies into account (Bolker et al., 2009; Harrison et al., 2018; see also Gurgel-Gonçalves et al., 2021). Although we built a unique collection of nearly 6600 digital pictures featuring 730 individual specimens, our species-level coverage was still limited; therefore, here we focus on algorithm performance at the genus level, for which we were able to gather a large enough sample of bugs and pictures. We follow the currently accepted systematic arrangement for the Triatominae (Monteiro et al., 2018), with all our bugs falling into one of the three genera to which the main vectors of human Chagas disease, from the USA to Argentina, all belong – *Triatoma*, *Panstrongylus*, and *Rhodnius* (WHO, 2002).

2. Methods

2.1. Bug-picture database

We worked with a sample of 730 adult, dried, pinned triatomines belonging in 13 species within three genera – *Triatoma* (400 individual specimens), *Panstrongylus* (110 specimens), and *Rhodnius* (220 specimens) (Table 1). RG-G and VLdM completed ‘ground-

truth' species-level identifications using printed keys and descriptions (Lent and Wygodzinsky, 1979; Galvão, 2014), the electronic TRIATODEX key (Gurgel-Gonçalves et al., 2021), and information on the geographical origin of the specimens. After removing the entomological pin, we placed each bug on a white, rotating circular base and pictured it nine times, one at each of nine different angles representing three positions – dorsal-flat, dorsal-oblique, and front/back-oblique (Fig. 1). All pictures were taken using the same Moto G6 Play cellphone (Motorola Mobility LLC, Chicago, IL), which was mounted on a tripod and positioned above the bug as shown in Fig. 1; we used neither the flash nor the zoom of the cellphone camera. We saved the pictures in red-green-blue (RGB) format at 72-dpi resolution, and cropped them into a square format (see Fig. 1). Overall, our dataset thus includes 6570 unique pictures featuring 730 individual bugs (Table 1). The pictures will be made available in the Dryad Digital Repository upon publication, and the full raw dataset, including picture metadata, is provided in Supplementary Dataset S1.

Table 1. Cellphone picture-based automated identification of Chagas disease vectors: triatomine bugs and pictures used in the study.

Genus	Species	Bug position in picture*			Total
		A	B	C	
<i>Triatoma</i>	<i>Triatoma brasiliensis</i>	130	780	260	1170
	<i>Triatoma costalimai</i>	40	240	80	360
	<i>Triatoma infestans</i>	60	360	120	540
	<i>Triatoma lenti</i>	40	240	80	360
	<i>Triatoma melanocephala</i>	20	120	40	180
	<i>Triatoma rubrovaria</i>	30	180	60	270
	<i>Triatoma sherlocki</i>	30	180	60	270
	<i>Triatoma sordida</i>	30	180	60	270
	<i>Triatoma vitticeps</i>	20	120	40	180
	Subtotal	400	2400	800	3600
<i>Panstrongylus</i>	<i>Panstrongylus lutzi</i>	20	120	40	180
	<i>Panstrongylus megistus</i>	90	540	180	810
	Subtotal	110	660	220	990
<i>Rhodnius</i>	<i>Rhodnius neglectus</i>	200	1200	400	1800
	<i>Rhodnius pictipes</i>	20	120	40	180
	Subtotal	220	1320	440	1980
Grand total		730	4380	1460	6570

*Bugs were pictured at nine angles representing three positions: dorsal-flat (**A**; one picture per bug, so that figures in this column also indicate the number of unique specimens in each class); dorsal-oblique (**B**; six pictures per bug); and front/back-oblique (**C**; two pictures per bug); see Fig. 1

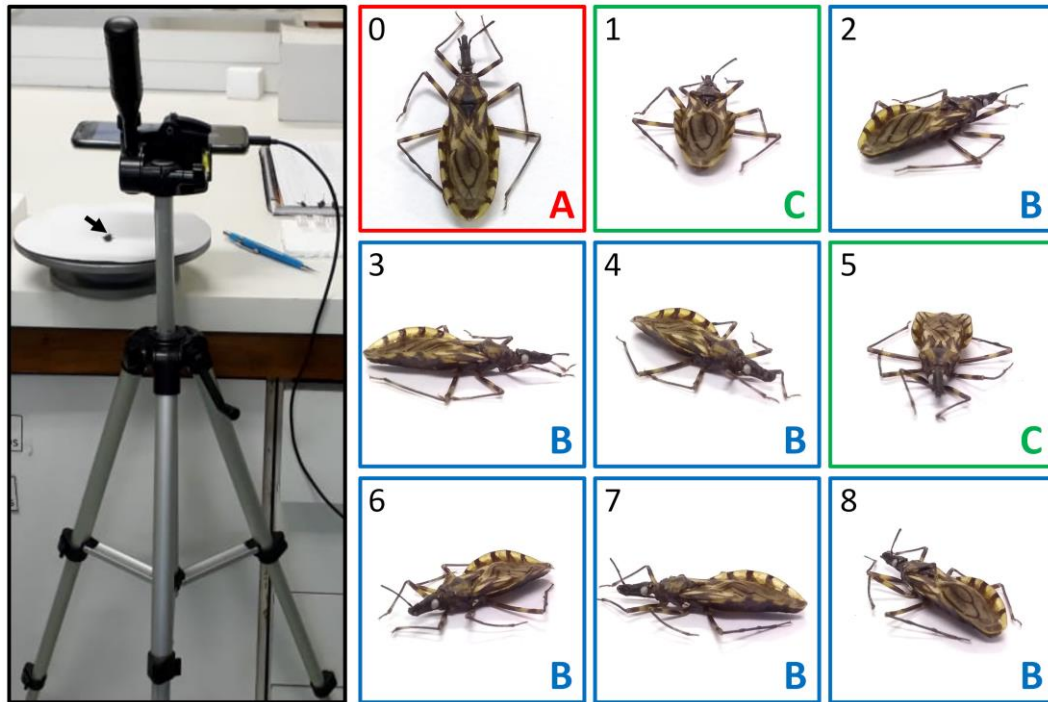


Fig. 1. Picturing bugs at different angles. Left panel: a tripod-mounted cellphone is positioned over a circular, rotating white base where the bug to be pictured (arrow) is placed. Right panels: each bug (here, a *Triatoma brasiliensis* specimen) was pictured at nine angles (0–8) representing three positions (colored letters and picture outlines): A–red, dorsal-flat (picture 0); B–blue, dorsal-oblique (pictures 2–4, 6–8); and C–green, front/back-oblique (pictures 1 and 5).

2.2. Algorithms

We used five machine-learning algorithms based on three different principles: a pre-trained convolutional neural network (AlexNet); three boosting-based classifiers (AdaBoost Adaptive Boosting, Gradient Boosting, and Histogram-based Gradient Boosting); and a Linear Discriminant model. Here, we briefly describe the main characteristics of these algorithms and the input they require; further details can be found in the references provided below.

The architecture of AlexNet (‘AN’ hereafter), which includes five convolutional layers and three fully-connected layers, has 60 million parameters; AN uses RGB images, 227×227 pixels in size, as the input, and was pre-trained on 1.2 million images of varying resolution

and representing 1000 classes (i.e., different objects, animals, plants, etc.) from the ImageNet database (Krizhevsky et al., 2012).

We also tested four ‘classical’ machine-learning algorithms – AdaBoost Multi-Class Adaptive Boosting (‘AB’ hereafter), Gradient Boosting (‘GB’ hereafter), Histogram-based Gradient Boosting (‘HB’ hereafter), and a Linear Discriminant model (‘LD’ hereafter). Instead of directly taking pictures as the input, these algorithms use numerical ‘features’ extracted from, and representing attributes of, those pictures. We used *Inovtaxon* (Bambil et al., 2020) to extract 226 picture features (namely, 36 color features, 135 shape features, and 55 texture features) that were used as input to feed the algorithms (see Bambil et al., 2020 and <https://github.com/DeborahBambil/Inovtaxon> for details; see also Hu, 1962; Zhao and Pietikainen, 2007; Flusser et al., 2009; Deniz et al., 2011; Nascimento et al., 2023).

AB combines multiple weak classifiers into a strong classifier capable of handling multi-class problems; at each training iteration, AB uses weights to ‘tag’ misidentified samples so that later iterations can ‘learn’ these mistakes, thus improving overall accuracy (Freund and Schapire, 1997). GB sequentially combines multiple decision trees, with iterative adjustment (‘learning’) based on the residual errors of previous models and minimization based on gradient descent (Friedman, 2001). HB is a fast, computationally efficient approach to gradient boosting in which continuous input variables (‘features’) are discretized into bins and then arranged as histograms; model training is thus made much faster and memory efficient, allowing for the use of gradient boosting with large datasets (Dalal and Triggs, 2005). Finally, LD is a classical dimension-reduction procedure in which linear functions of input values are computed that maximize among-class relative to within-class variance, thus maximizing class discrimination (Fisher, 1936). In the context of machine learning, these ‘linear discriminant functions’ are computed during training and then used to assign test instances to their most likely class (Tharwat et al., 2017).

2.3. Algorithm training and testing

We trained all algorithms on a random subset of 5256 pictures (80% of the 6570-picture dataset). To gauge performance consistency, the training-testing process was repeated 10 times for each algorithm; in each of these (pseudo-)replicate runs, all five algorithms used the same training and testing picture subsets.

AN was implemented in MATLAB (www.mathworks.com); for training, we used the stochastic gradient descent with momentum optimizer with default settings except for the initial learning rate (which we set to 0.001) and the maximum number of epochs (which we set to 15 after preliminary tests). The ‘classical’ machine-learning algorithms (AB, GB, HB, and LD) were run in Python (www.python.org) using the *scikit-learn* toolbox (Pedregosa et al., 2011). We left hyperparameters at *scikit-learn* default values, except for the learning rate of GB and HB (which we set to 0.15) and the tolerance threshold value for LD (set to 1×10^{-5}).

In each of the 10 (pseudo-)replicate runs, all algorithms were tested on the same subset of 1314 ‘problem pictures’ – for each of which we scored whether the class label predicted by the algorithm did/did not match the ‘ground-truth’ label. Correct/incorrect binary predictions were used as outcome variables in downstream analyses.

2.4. Data analysis: assessing algorithm performance

We analyzed the outcome of our 50 testing runs (10 per algorithm) using R 4.2.1 (R Core Team, 2022) and the RStudio 2023.03.1.446 interface (Posit Software, 2022). We first summarized our data in descriptive tables, including cross-classification or ‘confusion’ matrices, and ran simple exploratory analyses by calculating and graphing frequencies and proportions. Proportions and their Wilson ‘score’ 95% confidence intervals (CIs) (Newcombe, 1998) were computed using the *Hmisc* R package (Harrell and Dupont, 2023).

Next, we used generalized linear mixed models (GLMMs; Bolker et al., 2009; Harrison et al., 2018) with binomial error distribution and logit link-function to investigate (i) whether

and to what extent two key identification-performance metrics, accuracy and specificity, varied across algorithms ('algorithm' predictor, a five-level factor) and bug genera ('genus' predictor, a three-level factor), and (ii) the effects of cellphone-picture characteristics (picture 'angle', a nine-level factor; or bug 'position', a three-level factor; see Fig. 1) on algorithm performance. Covariates 'angle' and 'position', which are correlated by construction, did not appear together in any model. Initially, all models included a 'picture-within-bug' nested random effect to account for the non-independence of results arising from the same picture and from pictures of the same specimen; in some cases, however, this led to model convergence issues and the models were simplified (see *Specificity* below). Because in each (pseudo-)replicate run all algorithms ran on the same data (i.e., picture) subsets, we also included a 'replicate' random effect in the models.

We fit our GLMMs in a Bayesian framework using the *blme* R package, with weak, normal ($\mu = 0$, $\sigma^2 = 9$) fixed-effect priors and flat covariance priors (Chung et al., 2013, 2015; Bolker, 2018; Dorie et al., 2022). Relative model performance was evaluated based on sample-size corrected Akaike's information criterion (AICc) scores and related metrics (Burnham and Anderson, 2002) computed with the *bbmle* R package (Bolker et al., 2022). AICc calculations used the number of unique specimens in each dataset/subset (see below) as the sample size. For inference, we focus on the top-ranking (smallest-AICc) model in each model set (accuracy plus genus-specific specificity; see below) (Burnham and Anderson, 2002).

Accuracy. Accuracy analyses made use of the full, 65,700-observation dataset. The binary dependent variable specified whether each picture was (coded 1) or was not (coded 0) correctly identified in each identification task. Note that stratification by genus (i.e., inclusion of the 'genus' fixed effect in a model) implies conditioning on bug genus, so that genus-specific accuracy estimates can be interpreted as estimates of *sensitivity* – defined as the

probability of getting a correct answer, *conditioned on genus identity*. We fitted (1) an intercept-only model; (2) four bivariate models, each including one of the variables of interest ('algorithm'; bug 'genus'; picture 'angle'; bug 'position'); (3) five additive models including 'algorithm' and combinations of the other predictors; and (4) five interaction models including 'algorithm \times genus'; 'algorithm \times position'; and 'genus \times position' interaction terms (see Table 2). We initially also considered (i) two-way interactions involving the nine-level 'angle' covariate and (ii) three-way ('algorithm \times genus \times position') interactions; these highly complex models, however, often showed signs of convergence issues, and we did not consider them further.

Table 2. Accuracy model set. All models included a picture-within-bug nested random effect and a (pseudo-)replicate run random effect.

Model	AICc	Δ AICc	k	Weight	Terms (fixed effects)			
					Algorithm	Genus	Angle	Position
A×G+P	17246.0	0.0	20	0.87	×	×		
A×G	17249.8	3.8	18	0.13	×	×		
A+G×P	18783.0	1537.0	16	<0.001		×		×
A+G+P	18783.3	1537.3	12	<0.001				
A+G+An	18786.1	1540.1	18	<0.001				
A+G	18786.5	1540.5	10	<0.001				
A×P+G	18787.6	1541.6	20	<0.001	×			×
A+P	19260.6	2014.6	10	<0.001				
A+An	19262.7	2016.7	16	<0.001				
A	19263.7	2017.7	8	<0.001				
A×P	19264.3	2018.3	18	<0.001	×			×
G	31412.3	14166.3	6	<0.001				
P	31862.4	14616.4	6	<0.001				
Null	31864.7	14618.7	4	<0.001				
An	31865.2	14619.2	12	<0.001				

AICc, sample size-corrected Akaike’s information criterion; Δ AICc, AICc difference with the top-ranking (lowest-AICc) model; k , number of parameters; Weight, Akaike weight

Shaded cells indicate that the term was included (as a fixed effect) in the model; the ‘×’ symbol indicates which terms are involved in an interaction

Specificity. For a given bug genus, specificity analyses were based on the subset of data featuring only pictures of bugs *not* belonging in that genus. The binary dependent variable specified whether each picture was (coded 1) or was not (coded 0) correctly identified as *not* belonging in the focal genus, conditional on the picture featuring a bug indeed *not* belonging in that genus. Some of the specificity models had clear signs of convergence issues, most likely because correct-identification frequencies were very high across data subsets and, within them, across strata. We therefore simplified the structure of our specificity models by dropping the ‘angle’ fixed effect, interaction terms, and the ‘picture-within-bug’ random effect. Note that this simplification means that (i) subtler patterns of variation in specificity may exist that our models do not capture, and (ii) uncertainty around model coefficients (and, hence, around model predictions) may be underestimated. For each genus, the specificity-analysis model set thus included (1) an intercept-only model; (2) three bivariate models

(‘algorithm’; ‘genus’; bug ‘position’); and (3) three additive models (‘algorithm’ + ‘genus’; ‘algorithm’ + ‘position’; and ‘algorithm’ + ‘genus’ + ‘position’) (see Table 3).

Table 3. Specificity model sets. All models included a (pseudo-)replicate random effect. Note that there is one model set for each genus.

Genus	Model	AICc	Δ AICc	<i>k</i>	Weight	Terms (fixed effects)		
						Algorithm	Genus	Position
<i>Triatoma</i>	T A+G+P	14221.9	0.0	9	0.529	Shaded	Shaded	Shaded
	T A+G	14223.7	1.8	7	0.220	Shaded	Shaded	White
	T A+P	14224.1	2.2	8	0.177	White	White	Shaded
	T A	14225.8	3.9	6	0.074	Shaded	White	White
	T G	23904.8	9682.9	3	<0.001	White	Shaded	White
	T Null	23905.3	9683.4	2	<0.001	White	White	White
	T P	23905.8	9683.9	4	<0.001	White	White	Shaded
<i>Panstrongylus</i>	P A+G+P	7592.5	0.0	9	0.830	Shaded	Shaded	Shaded
	P A+G	7595.8	3.2	7	0.170	Shaded	Shaded	White
	P A+P	8477.6	885.1	8	<0.001	White	White	Shaded
	P A	8482.1	889.6	6	<0.001	Shaded	White	White
	P G	8778.5	1186.0	3	<0.001	White	Shaded	White
	P P	9643.5	2051.0	4	<0.001	White	White	Shaded
	P Null	9647.8	2055.3	2	<0.001	White	White	White
<i>Rhodnius</i>	R A+G+P	2327.2	0.0	9	0.992	Shaded	Shaded	Shaded
	R A+G	2336.9	9.7	7	0.008	Shaded	Shaded	White
	R A+P	2419.6	92.4	8	<0.001	White	White	Shaded
	R G	2429.1	101.9	3	<0.001	White	Shaded	White
	R A	2429.4	102.2	6	<0.001	Shaded	White	White
	R P	2511.5	184.3	4	<0.001	White	White	Shaded
	R Null	2521.5	194.3	2	<0.001	White	White	White

AICc, sample size-corrected Akaike’s information criterion; Δ AICc, AICc difference with the top-ranking (lowest-AICc) model; *k*, number of parameters; Weight, Akaike weight
Shaded cells indicate that the term was included (as a fixed effect) in the model; all models are additive

3. Results

3.1. Accuracy

Fig. 2 summarizes the main results of the full set of 50 testing runs. The pre-trained convolutional neural network, AN, correctly identified the genus of the specimen featured in each of the 5896 unique pictures it was tested on across the 10 (pseudo-)replicate runs. In contrast, AB overall accuracy was only ~0.692, with (pseudo-)replicate run-specific values

ranging from 0.595 to 0.760. Accuracy was high (~0.976 overall) for HB, and was also above 95% for GB (~0.965) and LD (~0.959); variation across (pseudo-)replicate runs was small for these three algorithms (Fig. 2). On average, genus-level identification was more accurate for pictures featuring *Triatoma* specimens (observed sensitivity 0.967) than for those featuring bugs in *Panstrongylus* (0.855) or *Rhodnius* (0.918) (Fig. 2). Finally, the data suggest that overall accuracy did not change much across picture angles or bug positions (Fig. 2). Detailed numerical results for these exploratory analyses of observed accuracy are provided in Supplementary Tables S1–S6, and the full raw data are available in Supplementary Dataset S1.



Fig. 2. Cellphone picture-based automated identification of Chagas disease vectors: observed accuracy, measured as the proportion of pictures that were correctly identified (blue bars) to the genus level. Left: results stratified by algorithm (AN, AlexNet; AB, AdaBoost Adaptive Boosting; GB, Gradient Boosting; HB, Histogram-based Gradient Boosting; and LD, Linear Discriminant model). Right, top row: results stratified by (i) triatomine-bug genus, which can be interpreted as sensitivity (the probability of correct identification, conditioned on genus); (ii) the nine angles at which bugs were pictured (see Fig. 1); and (iii) bug position (A, dorsal-flat; B, dorsal-oblique; C, front/back-oblique; see Fig. 1). Bottom-right: results for each of the 10 (pseudo-)replicate runs completed by each algorithm. The data underlying these graphs, along with score 95% confidence intervals for each proportion, are provided in Supplementary Tables S1–S5.

In Fig. 3 we present a set of ‘confusion’ matrices summarizing the patterns of correct/incorrect identification of bug pictures across algorithms and genera and over the 10 (pseudo-)replicate runs each algorithm completed. AB had a strong tendency to misidentify *Rhodnius* as *Triatoma*, and confusion between *Triatoma* and *Panstrongylus* was also evident. Although GB, HB, and especially LD also struggled somewhat to tell *Triatoma* from *Panstrongylus*, they did so to a much lesser extent than AB. On the other hand, there were no instances of confusion between *Panstrongylus* and *Rhodnius* (Fig. 3).

AlexNet (AN)		Prediction			Sum
		<i>Triatoma</i>	<i>Panstrongylus</i>	<i>Rhodnius</i>	
Truth	<i>Triatoma</i>	7193	0	0	7193
	<i>Panstrongylus</i>	0	1986	0	1986
	<i>Rhodnius</i>	0	0	3961	3961
Sum		7193	1986	3961	13,140
AdaBoost (AB)		Prediction			Sum
		<i>Triatoma</i>	<i>Panstrongylus</i>	<i>Rhodnius</i>	
Truth	<i>Triatoma</i>	6522	628	43	7193
	<i>Panstrongylus</i>	806	1180	0	1986
	<i>Rhodnius</i>	2577	0	1384	3961
Sum		9905	1808	1427	13,140
Gradient-based Boosting (GB)		Prediction			Sum
		<i>Triatoma</i>	<i>Panstrongylus</i>	<i>Rhodnius</i>	
Truth	<i>Triatoma</i>	7008	110	75	7193
	<i>Panstrongylus</i>	216	1770	0	1986
	<i>Rhodnius</i>	61	0	3900	3961
Sum		7285	1880	3975	13,140
Histogram-based Boosting (HB)		Prediction			Sum
		<i>Triatoma</i>	<i>Panstrongylus</i>	<i>Rhodnius</i>	
Truth	<i>Triatoma</i>	7086	68	39	7193
	<i>Panstrongylus</i>	165	1821	0	1986
	<i>Rhodnius</i>	43	0	3918	3961
Sum		7294	1889	3957	13,140
Linear Discriminant model (LD)		Prediction			Sum
		<i>Triatoma</i>	<i>Panstrongylus</i>	<i>Rhodnius</i>	
Truth	<i>Triatoma</i>	6982	173	38	7193
	<i>Panstrongylus</i>	256	1730	0	1986
	<i>Rhodnius</i>	73	0	3888	3961
Sum		7311	1903	3926	13,140

Fig. 3. Cellphone picture-based automated identification of Chagas disease vectors: ‘confusion’ (cross-classification) matrices across algorithms and bug genera. The numbers of pictures that were correctly/incorrectly identified at the genus level are in blue/red cells, respectively. Note that, for each algorithm, the results arise from 10 (pseudo-)replicate runs; in each run, all algorithms were trained and tested on the same picture subsets, which were randomly drawn from the full, 6570-picture dataset.

The top-performing accuracy GLMM had strong relative support from the data, with an AICc score 3.8 units smaller than that of the second-ranking model and an Akaike weight of 0.87 (Table 2). This top-ranking model includes an ‘algorithm × genus’ interaction term and an additive bug ‘position’ term as fixed effects. In Fig. 4 we provide a plot of model predictions for each genus, algorithm, and bug position (see Supplementary Table S7 for numerical model output). The model captures the essentially perfect accuracy of AN, and suggests that variation across algorithms was contingent on bug genus – with, e.g., AB performing fairly well with *Triatoma*, yet poorly with *Rhodnius* and *Panstrongylus*, and both GB and LD struggling somewhat with *Panstrongylus* (Fig. 4, Supplementary Table S7). In

contrast, bug-position effects were relatively small; if anything, dorsal-flat pictures perhaps improved accuracy, albeit only slightly, relative to dorsal-oblique pictures (Fig. 4). The top-ranking model estimated a much smaller random-effect variance for the (pseudo-)replicate grouping factor ($\sigma^2 = 0.065$) than for the specimen ($\sigma^2 = 3.683$) or picture ($\sigma^2 = 2.546$) grouping factors (Supplementary Table S7). The second-ranking model (Akaike weight = 0.13; Table 2) does not include the bug ‘position’ covariate – whose effect, therefore, is estimated to be effectively zero by that model. Other models, including the ‘null’ model, had no support from the data (Table 2). We finally note that models including the nine-level ‘angle’ covariate performed consistently worse (i.e., had larger AICc scores) than their counterparts including the simpler, four-level ‘position’ covariate (Table 2). In line with the stratification-by-angle results shown in Fig. 2 and Supplementary Table S3, this suggests that picture angle had negligible effects on accuracy.

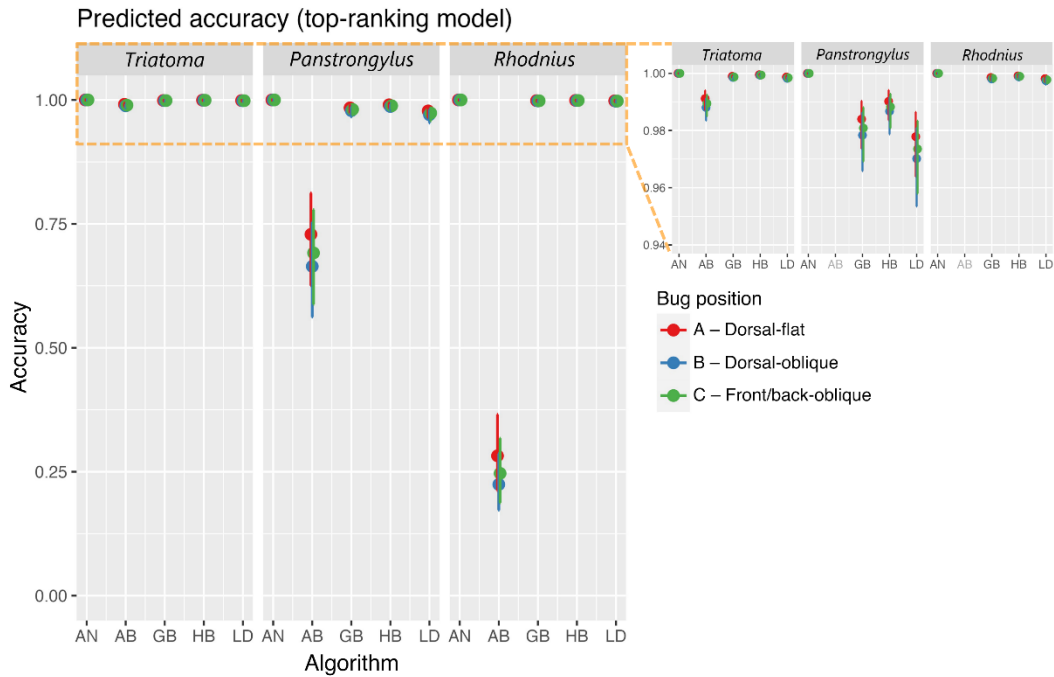


Fig. 4. Cellphone picture-based automated identification of Chagas disease vectors: predictions (with 95% confidence intervals) of the top-ranking generalized linear mixed model for accuracy (see Table 2 and Supplementary Table S7). Accuracy predictions are presented for five algorithms (AN, AlexNet; AB, AdaBoost Adaptive Boosting; GB, Gradient Boosting; HB, Histogram-based Gradient Boosting; and LD, Linear Discriminant model), three triatomine-bug genera (*Triatoma*, *Panstrongylus*, and *Rhodnius*), and three bug positions (see legend and Fig. 1). The top-right panel is a zoom-in view of the region highlighted by a dashed-orange box in the main graph. The data underlying this figure are provided in Supplementary Table S8.

3.2. Specificity

Specificity measures the complement of the false-positive rate, providing insight into misclassification probabilities. As noted above, AN consistently identified all bugs in their correct genus, so that specificity was also essentially perfect (Fig. 5; see also Fig. 3). On the other extreme, most of the bugs identified as ‘*Triatoma*’ by AB were not *Triatoma*, but either *Panstrongylus* or *Rhodnius*, yielding a low observed specificity of ~0.431 (Fig. 5; see also Fig. 3 and Supplementary Table S9). Identifications of ‘*Triatoma*’ by GB, HB, and LD were overall more reliable, yet observed specificities were modest, from ~0.945 for LD to ~0.965 for HB (Figs. 3 and 5; Supplementary Table S9). Except for AB (observed specificity ~0.944), non-*Panstrongylus* bugs were rarely confused with *Panstrongylus* by the remaining algorithms (Figs. 3 and 5). Finally, only a very small fraction of pictures featuring non-

Rhodnius bugs were misidentified as featuring a *Rhodnius* specimen, with specificities consistently above 99% (Figs. 3 and 5; Supplementary Table S9).

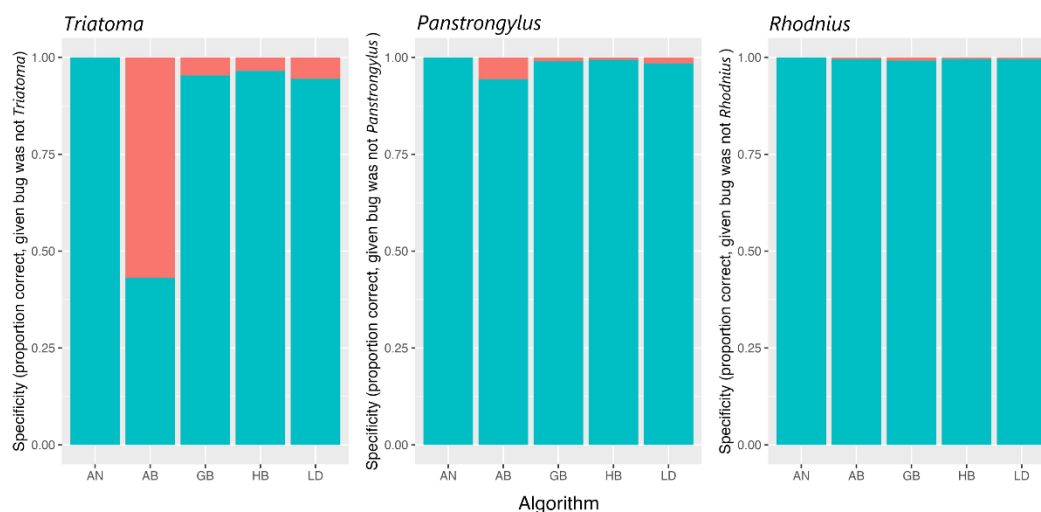


Fig. 5. Cellphone picture-based automated identification of Chagas disease vectors: observed specificity, measured as the proportion of pictures that were correctly identified (blue bars) as *not* belonging in the genus given above each graph. Note that, in this case, red bars represent the false-positive rate. For each genus, results were stratified by algorithm (AN, AlexNet; AB, AdaBoost Adaptive Boosting; GB, Gradient Boosting; HB, Histogram-based Gradient Boosting; and LD, Linear Discriminant model). The data underlying these graphs, along with score 95% confidence intervals for each proportion, are provided in Supplementary Table S9.

Table 3 summarizes the structure and relative performance of the seven models in each of the three model sets we fit to assess specificity. In all cases, the top-ranking model was the ‘full’ additive model including the three covariates we considered (‘algorithm’, ‘genus’, and ‘position’), and all data-supported models consistently included the ‘algorithm’ covariate (Table 3). For *Triatoma*, the second- and third-ranking models had also some support from the data (Table 3); they differ from the top-ranking model in that they lack, respectively, the ‘position’ and ‘genus’ covariates, suggesting that their effects were likely small. The numerical output of the top-ranking model confirmed this, with the absolute values of coefficient estimates ranging from 0.096 to 0.168 (Supplementary Table S10). Similarly, the second-ranking *Panstrongylus* specificity model lacks the ‘position’ covariate, again

indicating that it had, if anything, small effects; in the top-ranking model, the coefficients were small and their 95% CIs included 0 (see Table 3 and Supplementary Table S11). In contrast, the top-ranking *Rhodnius* model was strongly supported (Akaike weight = 0.99; Table 3), and its numerical output suggested a clearer negative effect of dorsal-oblique pictures on specificity ($\beta_{\text{Position B}} = -0.594$, SE 0.277; Supplementary Table S12).

In Fig. 6 we present the predictions of the three top-ranking specificity models. The graphs highlight (i) the essentially perfect performance of AN, (ii) the very poor performance of AB with non-*Triatoma* bug pictures, and (iii) the overall high to very high, often close to 100%, specificity of all algorithms with non-*Panstrongylus* and non-*Rhodnius* bugs (Fig. 6; see also Supplementary Tables S13– S15). We recall, however, that these models do not include picture or specimen random effects (which led to convergence issues); in interpreting model-based estimates and predictions, therefore, it is important to keep in mind that our measures of uncertainty about them may be smaller than they should.

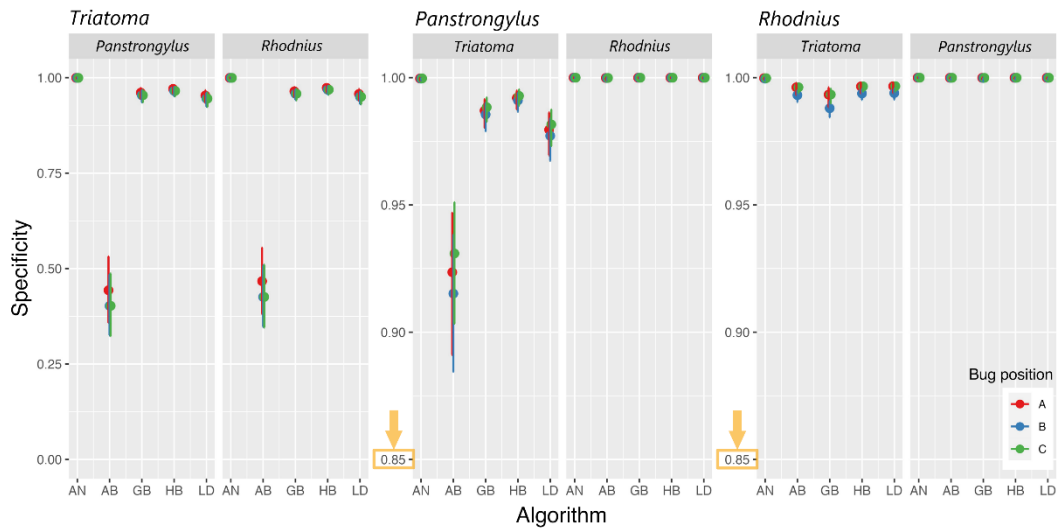


Fig. 6. Cellphone picture-based automated identification of Chagas disease vectors: predictions (with 95% confidence intervals) of the top-ranking generalized linear mixed models for specificity (see Table 3 and Supplementary Tables S10–S12). Predictions are presented for three triatomine-bug genera (*Triatoma*, *Panstrongylus*, and *Rhodnius*), five algorithms (AN, AlexNet; AB, AdaBoost Adaptive Boosting; GB, Gradient Boosting; HB, Histogram-based Gradient Boosting; and LD, Linear Discriminant model), and three bug positions (A, dorsal-flat; B, dorsal-oblique; C, front/back oblique; see Fig. 1). The orange boxes and arrows highlight the different scales of the y-axes (starting at 0.85 instead of at 0.00) in the center and right panels. The data underlying this figure are provided in Supplementary Tables S13–S15.

4. Discussion

In this study we show that deep-learning algorithms can use low-resolution pictures taken at varying angles with ordinary cellphone cameras to reliably identify Chagas disease vectors at the genus level. This brings us a step closer to the broader aim of developing automated vector-identification tools that, by leveraging information from potentially large, geographically diffuse networks of cellphone users (including vector control-surveillance staff and citizens in general), can help strengthen real-world vector-surveillance systems in the current context of dwindling taxonomic expertise (Gurgel-Gonçalves, 2022). More generally, our results contribute to continued efforts towards structuring ‘extended’ surveillance systems in which classical and novel approaches to vector detection, identification, reporting, and mapping can be integrated into more effective, flexible tools for risk assessment and decision making (e.g., Marsden, 1984; Silveira et al., 1984; Gorla, 2002;

Abad-Franch et al., 2009, 2011, 2013, 2014; Leite et al., 2011; Gurgel-Gonçalves et al., 2012, 2017, 2021; Vinhaes et al., 2014; Curtis-Robles et al., 2015; Abad-Franch, 2016; Oliveira et al., 2017; Ceccarelli et al., 2018, 2020, 2022; Hamer et al., 2018; Khalighifar et al., 2019; Bender et al., 2020; Cruz et al., 2020, 2021; Abdelghani et al., 2021; Gürtler and Cecere, 2021; Parsons and Banitaan, 2021; Ribeiro-Jr et al., 2021; Cochero et al., 2022; Delgado-Noguera et al., 2022).

Our comparative assessment singled out the pre-trained, convolutional neural network, AlexNet (Krizhevsky et al., 2012) as clearly the best-performing algorithm among the five we tested (Figs. 2–6). This is in line with previous work showing that artificial neural networks can be from fairly to very accurate at identifying disease vectors (Motta et al., 2019; Park et al., 2020; Pataki et al., 2021; Cochero et al., 2022) and other insects (e.g., Terry et al., 2020) when trained and tested on non-standardized pictures sourced from field research or citizen-science initiatives. When trained and tested on better-quality pictures (e.g., high-resolution pictures of well-lit whole specimens), artificial neural networks can be even more accurate at identifying, for example, triatomine bugs (Gurgel-Gonçalves et al., 2017; Khalighifar et al., 2019; Abdelghani et al., 2021), mosquitoes (Motta et al., 2020), or ticks (Justen et al., 2021). Because real-world ‘extended’ surveillance systems will have to rely on pictures of varying quality, most likely taken with cellphone cameras, automated-identification systems that perform well when trained and tested on low-resolution, non-standardized pictures will be critical to advancing the field; our results lend further support to the view that pre-trained artificial neural networks are strong candidates for the role (Motta et al., 2019; Justen et al., 2021; Pataki et al., 2021; Cochero et al., 2022).

It is important to note that the consistently perfect performance of AN we report (Figs. 2, 3; Supplementary Tables S1, S5, and S6) refers to a relatively simple, three-class (i.e., three-genus) problem. For any given algorithm and picture quality, identification tends to

become more challenging as the number of classes increases – e.g., when the task involves several classes (e.g., Khalighifar et al., 2019) instead of just two (e.g., Cochero et al., 2022) or three classes (e.g., Cruz et al. 2021 and this report). In our previous work, the accuracy of two artificial neural networks rose from ~80–87% to ~96–99% when the number of classes was reduced from 12–39 species to subsets of 2–12 candidate species – with subsetting based on species-specific distribution records and models (Khalighifar et al., 2019). While this ‘faunal subsetting’ approach may help improve the accuracy of automated-identification systems, defining such subsets based on (known or modeled) species distributions is fraught with potential pitfalls. Most obviously, species-occurrence records are incomplete and may be biased because of, e.g., uneven sampling effort or false-positive records (Ceccarelli et al., 2018, 2022; Johnston et al., 2023); perhaps less obviously, the methodological challenges inherent to modeling presence records (which are likely incomplete and possibly biased) mean that model-predicted species distributions should be approached with caution (Elith et al., 2006; Owens et al., 2013; Hallgren et al., 2019; Araujo et al., 2019; Valavi et al., 2022). Also importantly, humans have historically introduced several triatomine-bug species well beyond their natural ranges (Monteiro et al., 2018). Because introductions may go undetected for long periods, particularly when surveillance is weak (Abad-Franch et al., 2014), defining a realistic subset of the triatomine fauna likely to be present at any given location and time-point may be far from easy; notably, the procedure will by default exclude non-native species, which are those better adapted to living in human habitats, hence the most dangerous vectors of human Chagas disease, and hence the top priority for control-surveillance (Abad-Franch, 2016; Gürtler et al., 2021). Careful ‘faunal subsetting’ based on the best available evidence, in any case, clearly has a major role to play in both ‘classical’ (e.g., Ribeiro-Jr et al., 2021) and ‘extended’ vector surveillance – in which accurate automated identification of the 150+ known triatomine-bug species, some of which form cryptic-species complexes and some of

which are highly variable (Monteiro et al., 2018), may well prove unfeasible without some kind of subsetting (Khalighifar et al., 2019). In general, the use of reliable specimen-specific metadata (on, e.g., collection sites and dates) can be expected to improve the accuracy of automated identification systems (Terry et al., 2020).

Our report also illustrates a mixed model-based approach to evaluating the performance of automated identification systems while taking data dependencies into account (Bolker et al., 2009; Harrison et al., 2018; see also Gurgel-Gonçalves et al., 2021). In our case and in many similar studies, a large number of observations (here, the individual outcomes of 65,700 identification tasks) arose from repeatedly using a much smaller number of unique pictures (here, 6570) featuring a much smaller number of unique specimens (here, 730 bugs), often across several training-testing runs (here, 10 runs) completed by several algorithms – five, in our case. Thus, for example, we used picture ‘Plutzix10y0’ (a dorsal-flat picture of a *Panstrongylus lutzi* specimen) 15 times (in three (pseudo-)replicate runs), and the bug featured in that picture (‘Plutzix10’) contributed 85 observations (over all (pseudo-)replicates) to the full dataset (see Supplementary Dataset S1). These 85 results cannot be treated as mutually independent, and the random-effects structure of our accuracy models takes the underlying dependencies into account (Harrison et al., 2018). GLMMs also provide insight into the magnitude of random variation across grouping factors; in our case, for example, accuracy was much more consistent across (pseudo-)replicate runs than it was across specimens or pictures (see Supplementary Table S9). While this analytical approach is a strength of our report, it also highlights a limitation of some of our models. In particular, convergence issues (likely related to very high correct-identification frequencies across strata) forced us to drop the ‘picture-within-bug’ nested random effect in specificity models (see Table 3, Fig 5, and Supplementary Table S9). As explained in the Methods section and

mentioned in the Results section, this simplification means that the output of those models must be interpreted with some caution.

We finally note that our Bayesian approach to model fitting relied on the selection of what we regard as reasonable prior distributions for model parameters – weakly informative $N(0, 9)$ priors for the fixed effects and flat covariance priors for the random effects (Bolker, 2018; Dorie et al., 2022). To see if our conclusions were robust to these choices, we refit each of the four top-ranking models (Tables 2, 3, Figs. 4 and 6, and Supplementary Tables S7, S8, and S10–S15) with the following alternative prior specifications: (i) fixed effects: $N(0, 14)$ and $N(0, 2.5)$; and (ii) random effects: Wishart (the default in *blme*; Dorie et al., 2022; see also Chung et al., 2015), $\text{gamma}(\text{shape}=2.5, \text{rate}=0)$ (the default gamma specification in *blme*; Dorie et al., 2022), and $\text{gamma}(\text{shape}=2.5, \text{rate}=0.5)$. The results were highly robust, with only slight numerical differences, across all comparisons (see Supplementary Figs. S1–S4).

5. Conclusions and outlook

Automated identification of Chagas disease vectors has the potential to strengthen entomological surveillance systems – which are currently weak, and overall getting weaker, across Latin America. To fully realize such potential, automated-identification algorithms should be capable of working accurately with pictures of uneven, possibly low quality/resolution taken with ordinary cellphone cameras. Our results suggest that convolutional neural networks may indeed have that capacity – and, hence, that human expertise and artificial intelligence could team up to provide fast, reliable taxonomic support to ‘extended’, participatory surveillance systems with broad geographic-ecological scope. However, not all triatomine-bug species have the same public-health relevance, and future research should focus on combining mixed-quality pictures and pre-trained neural networks to reliably identify adult triatomine bugs to the species level. Developing automated systems

capable of telling triatomine-bug nymphs (i.e., immature stages, whose presence in or around houses often signals higher risk of disease transmission) from adults and from the nymphs of other, non-blood-feeding bugs, should be seen as another priority. A crucial standing challenge to the advancement of these prospects is the need to build ‘ground-truth’ picture datasets with adequate taxonomic coverage and large enough numbers of individual specimens (both nymphs and adults) to robustly train, validate, and test candidate algorithms. We anticipate that, when fully functional, ‘extended’ surveillance networks in which well-trained professional staff, mindful citizens, and efficient artificial-intelligence systems all contribute to vector detection, identification, reporting, and mapping will play critical roles in the sustainable, long-term control of vector-borne diseases including Chagas disease.

Author contributions

Vinicius Lima de Miranda: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – Original draft preparation. **Ewerton Pacheco de Souza:** Software, Methodology, Investigation, Writing – Review & editing. **Deborah Bambil:** Software, Methodology, Resources, Investigation, Writing – Review & editing. **Ali Khalighifar:** Conceptualization, Methodology, Writing – Review & Editing. **Andrew Townsend Peterson:** Conceptualization, Methodology, Writing – Review & Editing. **Francisco Assis de Oliveira Nascimento:** Conceptualization, Software, Methodology, Resources, Writing – Review & editing. **Rodrigo Gurgel-Gonçalves:** Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Supervision, Writing – Original draft preparation. **Fernando Abad-Franch:** Conceptualization, Data curation, Methodology, Formal analysis, Visualization, Validation, Supervision, Writing – Original draft preparation, Writing – Review & editing.

Declaration of competing interests

We have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Pictures will be made available in the Dryad Digital Repository upon publication; the full raw dataset, including picture metadata, is provided in Supplementary Dataset S1.

Acknowledgements

We thank José Roberto Pujol Luz (Universidade de Brasília), Antônio José Camilo Aguiar (Universidade de Brasília), Rita de Cássia Moreira de Souza (Fiocruz Minas Gerais), and Marcos Takashi Obara (Universidade de Brasília) for useful suggestions. The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Ministério da Educação, Brazil, provides support to the Programa de Pós-graduação em Zoologia at the Universidade de Brasília, Brazil.

References

- Abad-Franch, F., 2016. A simple, biologically sound, and potentially useful working classification of Chagas disease vectors. *Mem. Inst. Oswaldo Cruz* 111, 649–651.
- Abad-Franch, F., Diotaiuti, L., Gurgel-Gonçalves, R., Gürtler, R.E., 2013. Certifying the interruption of Chagas disease transmission by native vectors: *Cui bono?* *Mem. Inst. Oswaldo Cruz* 108, 251–254.
- Abad-Franch, F., Diotaiuti, L., Gurgel-Gonçalves, R., Gürtler, R.E., 2014. Reply. On bugs and bias: Improving Chagas disease control assessment. *Mem. Inst. Oswaldo Cruz*. 109, 125–130.

- Abad-Franch, F., Monteiro, F.A., Jaramillo, N., Gurgel-Gonçalves, R., Dias, F.B.S., Diotaiuti, L., 2009. Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: A multi-scale appraisal of the tribe Rhodniini (Triatominae). *Acta Trop.* 110, 159–177.
- Abad-Franch, F., Vega, M.C., Rolon, M.S., Santos, W.S., Rojas de Arias, A., 2011. Community participation in Chagas disease vector surveillance: Systematic review. *PLoS Negl. Trop. Dis.* 5, e1207.
- Abdelghani, B.A., Banitaan, S., Maleki, M., Mazen, A., 2021. Kissing bugs identification using convolutional neural network. *IEEE Access* 9, 140539–140548.
- Almeida, A.P.G., Fouque, F., Launois, P., Sousa, C.A., Silveira, H., 2017. From the laboratory to the field: Updating capacity building in medical entomology. *Trends Parasitol.* 33, 664–668.
- Araujo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5, eaat4858.
- Bambil, D., Pistori, H., Bao, F., Weber, V., Alves, F.M., Gonçalves, E.G., Figueiredo, L.F.A., Abreu, U.G.P., Arruda, R., Bortolotto, I.M., 2020. Plant species identification using color learning resources, shape, texture, through machine learning and artificial neural networks. *Environ. Syst. Decis.* 40, 480–484.
- Bender, A., Python, A., Lindsay, S.W., Golding, N., Moyes, C.L., 2020. Modelling geospatial distributions of the triatomine vectors of *Trypanosoma cruzi* in Latin America. *PLoS Negl. Trop. Dis.* 14, e0008411.
- Bolker, B.M., 2018. GLMM worked examples. Digression: Complete separation. https://bbolker.github.io/mixedmodels-misc/ecostats_chap.html#digression-complete-separation (accessed 19 July 2023).

- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.S.S., 2009. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* 24, 127–135.
- Bolker, B.M., R Development Team, Gine-Vazquez, I., 2022. Package ‘bbmle’. Tools for general maximum likelihood estimation. <https://cran.r-project.org/web/packages/bbmle/bbmle.pdf> (accessed 19 July 2023).
- Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer-Verlag, Berlin.
- Casas, J., Lazzari, C., Insausti, T., Launois, P., Fouque, F., 2016. Mapping of courses on vector biology and vector-borne diseases systems: Time for a worldwide effort. *Mem. Inst. Oswaldo Cruz* 111, 717–719.
- Ceccarelli, S., Balsalobre, A., Cano, M.E., Canale, D., Lobbia, P., Stariolo, R., Rabinovich, J.E., Marti, G.A., 2020. Analysis of Chagas disease vectors occurrence data: The Argentinean triatomine species database. *Biodiv. Data J.* 8, e58076.
- Ceccarelli, S., Balsalobre, A., Medone, P., Cano, M.E., Gurgel-Gonçalves, R., Feliciangeli, D., Vezzani, D., Wisnivesky-Colli, C., Gorla, D.E., Marti, G.A., Rabinovich, J.E., 2018. DataTri, a database of American triatomine species occurrence. *Sci. Data* 24, 180071.
- Ceccarelli, S., Balsalobre, A., Vicente, M.E., Curtis-Robles, R., Hamer, S.A., Landa, J.M.A., Rabinovich, J.E., Marti, G.A. (2022). American triatomine species occurrences: Updates and novelties in the DataTri database. *Gigabyte* 2022, gigabyte62.
- Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., Dorie, V., 2015. Weakly informative prior for point estimation of covariance matrices in hierarchical models. *J. Educ. Behav. Stat.* 40, 136–157.

- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., Liu, J., 2013. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* 78, 685–709.
- Cochero, J., Pattori, L., Balsalobre, A., Ceccarelli, S., Marti, G., 2022. A convolutional neural network to recognize Chagas disease vectors using mobile phone images. *Ecol. Inform.* 68, 101587.
- Cruz, D.D., Arellano, E., Denis Avila, D., Ibarra-Cerdeña, C.N., 2020. Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: A case for the cryptic *dimidiata* complex. *Parasit. Vectors* 13, 332.
- Cruz, D.D., Denis, D., Arellano, E., Ibarra-Cerdeña, C.N., 2021. Quantitative imagery analysis of spot patterns for the three-haplogroup classification of *Triatoma dimidiata* (Latreille, 1811) (Hemiptera: Reduviidae), an important vector of Chagas disease. *Parasit. Vectors* 14, 90.
- Curtis-Robles, R., Wozniak, E.J., Auckland, L.D., Hamer, G.L., Hamer, S.A., 2015. Combining public health education and disease ecology research: Using citizen science to assess Chagas disease entomological risk in Texas. *PLoS Negl. Trop. Dis.* 9, e0004235.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA 1, 886–893.
- Delgado-Noguera, L.A., Hernandez-Pereira, C.E., Ramirez, J.D., Hernandez, C., Velasquez-Ortiz, N., Clavijo, J., Ayala, J.M., Forero-Peña, D., Marquez, M., Suarez, M.J., Traviezo-Valles, L., Escalona, M.A., Perez-Garcia, L., Carpio, I.M., Sordillo, E.M., Grillet, M.E., Llewellyn, M.S., Gabaldón, J.C., Mondolfi, A.E.P., 2022. Tele-entomology and tele-parasitology: A citizen science-based approach for surveillance and control of Chagas disease in Venezuela. *Parasite Epidemiol. Control* 19, e00273.

- Deniz, O., Bueno, G., Salido, J., de la Torre, F., 2011. Face recognition using histograms of oriented gradients. *Pattern Recognit. Lett.* 32, 1598–1603.
- Dias, J.C.P., Ramos Jr, A.N., Gontijo, E.D., Luquetti, A., Shikanai-Yasuda, M.A., Coura, J.R., Torres, R.M., Melo, J.R.C., Almeida, E.A., de Oliveira Jr, W., Silveira, A.C., de Rezende, J.M., Pinto, F.S., Ferreira, A.W., Rassi, A., Fragata Filho, A.A., de Sousa, A.S., Correia Filho, D., Jansen, A.M., Andrade, G.M.Q., Britto, C.F.P., Pinto, A.Y.N., Rassi Jr, A., Campos, D.E., Abad-Franch, F., Santos, S.E., Chiari, E., Hasslocher-Moreno, A.M., Moreira, E.F., Marques, D.S.O., Silva, E.L., Marin-Neto, J.A., Galvão, L.M.C., Xavier, S.S., Valente, S.A.S., Carvalho, N.B., Cardoso, A.V., Silva, R.A., da Costa, V.M., Vivaldini, S.M., Oliveira, S.M., Valente, V.C., Lima, M.M., Alves, R., 2016. 2nd Brazilian Consensus on Chagas disease, 2015. *Rev. Soc. Bras. Med. Trop.* 49(Suppl. 1), 3–60.
- Dorie, V., Bates, D., Maechler, M., Bolker, B., Walker, S., 2022. Package ‘blme’. Bayesian linear mixed-effects models. <https://cran.r-project.org/web/packages/blme/blme.pdf> (accessed 19 July 2023).
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography* 29, 129–151.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, 179–188.
- Flusser, J., Zitova, B., Suk, T., 2009. Moments and Moment Invariants in Pattern Recognition. Wiley, New York.

- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput. Syst. Sci.* 55, 119–139.
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- GBD – Global Burden of Disease Collaborative Network, 2019. Global Burden of Disease Study 2019 (GBD 2019) results. Institute for Health Metrics and Evaluation, Seattle. <http://ghdx.healthdata.org/gbd-results-tool> (accessed 5 July 2023).
- Galvão, C. (Org.), 2014. Vetores da Doença de Chagas no Brasil. Sociedade Brasileira de Zoologia, Curitiba.
- Gorla, D.E., 2002. Variables ambientales registradas por sensores remotos como indicadores de la distribución geográfica de *Triatoma infestans* (Heteroptera: Reduviidae). *Ecol. Austral* 12, 117–127.
- Gurgel-Gonçalves, R., 2022. Stronger control-surveillance systems for vector-borne Chagas disease. *Mem. Inst. Oswaldo Cruz* 117, e210130chgsb.
- Gurgel-Gonçalves, R., Abad-Franch, F., de Almeida, M.R., Obara, M.T., de Souza, R.C.M., Batista, J.A.D.S., Rocha, D.D.A., 2021. TRIATODEX, an electronic identification key to the Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease: Development, description, and performance. *PLoS One*, 16, e0248628.
- Gurgel-Gonçalves, R., Galvão, C., Costa, J., Peterson, A.T., 2012. Geographic distribution of Chagas disease vectors in Brazil based on ecological niche modeling. *J. Trop. Med.* 2012, 705326.
- Gurgel-Gonçalves, R., Komp, E., Campbell, L.P., Khalighifar, A., Mellenbruch, J., Mendonça, V.J., Owens, H.L., Felix, K.C., Peterson, A.T., Ramsey, J.M., 2017. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: The Virtual Vector Lab. *PeerJ* 5, e3040.

- Gürtler, R.E., Cecere, M.C., 2021. Chagas disease vector control, in: Guarneri, A.A., Lorenzo, M.G. (Eds.), *Triatominae: The Biology of Chagas Disease Vectors*. Entomology in Focus, vol. 5. Springer, Cham, pp. 491–535.
- Gürtler, R.E., Fernández, M.P., Cardinal, M.V., 2021. Eco-epidemiology of vector-borne transmission of *Trypanosoma cruzi* in domestic habitats, in: Guarneri, A.A., Lorenzo, M.G. (Eds.), *Triatominae: The Biology of Chagas Disease Vectors*. Entomology in Focus, vol. 5. Springer, Cham, pp. 447–489.
- Hallgren, W., Santana, F., Low-Choy, S., Zhao, Y., Mackey, B., 2019. Species distribution models can be highly sensitive to algorithm configuration. *Ecol. Model.* 408, 108719.
- Hamer, S.A., Curtis-Robles, R., Hamer, G.L., 2018. Contributions of citizen scientists to arthropod vector data in the age of digital epidemiology. *Curr. Opin. Insect Sci.* 28, 98–104.
- Harrel, F., Dupont, C., 2023. Package ‘Hmisc’: Harrell miscellaneous. <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf> (accessed 19 July 2023).
- Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E., Robinson, B.S., Hodgson, D.J., Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6, e4794.
- Hashimoto, K., Zuniga, C., Romero, E., Morales, Z., Maguire, J.H., 2015. Determinants of health service responsiveness in community-based vector surveillance for Chagas disease in Guatemala, El Salvador, and Honduras. *PLoS Negl. Trop. Dis.* 9, e0003974.
- Hu, M.-K., 1962. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory*, 8, 179–187.
- Jansen, A.M., Roque, A.L.R., 2010. Domestic and wild mammalian reservoirs, in: Telleria, J., Tibayrenc, M. (Eds.), *American Trypanosomiasis (Chagas Disease): One Hundred Years of Research*. Elsevier, London, pp. 249–276.

- Johnston, A., Matechou, E., Dennis, E.B., 2023. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods Ecol. Evol.* 14, 103–116.
- Justen, L., Carlsmith, D., Paskewitz, S.M., Bartholomay, L.C., Bron, G.M., 2021. Identification of public submitted tick images: A neural network approach. *PLoS One* 16, e0260622.
- Khalighifar, A., Komp, E., Ramsey, J.M., Gurgel-Gonçalves, R., Peterson, A.T., 2019. Deep learning algorithms improve automated identification of Chagas disease vectors. *J. Med. Entomol.* 56, 1404–1410.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Sys.* 25, 1106–1114.
- Leite, G.R., dos Santos, C.B., Falqueto, A., 2011. Influence of the landscape on dispersal of sylvatic triatomines to anthropic habitats in the Atlantic Forest. *J. Biogeogr.* 38, 651–663.
- Lent, H., Wygodzinsky, P., 1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bull. Am. Mus. Nat. Hist.* 163, 123–520.
- Marsden, P.D., 1984. Selective primary health care: Strategies for control of disease in the developing world. XVI. Chagas' disease. *Rev. Infect. Dis.* 6, 855–865.
- Monteiro, F.A., Weirauch, C., Felix, M., Lazoski, C., Abad-Franch, F., 2018. Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. *Adv. Parasitol.* 99, 265–344.
- Motta, D., Santos, A.A.B., Machado, B.A.S., Ribeiro-Filho, O.G.V., Camargo, L.O.A., Valdenegro-Toro, M.A., Kirchner, F., Badaro, R., 2020. Optimization of convolutional neural network hyperparameters for automatic classification of adult mosquitoes. *PLoS One* 15, e0234959.

- Motta, D., Santos, A.A.B., Winkler, I., Machado, B.A.S., Pereira, D.A.D.I., Cavalcanti, A.M., Fonseca, E.O.L., Kirchner, F., Badaro, R., 2019. Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS One* 14, e0210829.
- Nascimento, F.A.O., Saraiva, R.G., Faria, E.G., Silva, T.A., Carvalho, J.L., 2023. Computational intelligence conceptions to automated diagnosis: Feature grouping for performance improvement. *Braz. Arch. Biol. Technol.* (in press).
- Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* 17, 857–872.
- Oliveira, M.L., Brito, R.N., Guimarães, P.A.S., dos Santos, R.V.M.A., Diotaiuti, L.G., Souza, R.C.M., Ruiz, J.C., 2017. TriatoKey: A web and mobile tool for biodiversity identification of Brazilian triatomine species. *Database* 2017, bax033.
- Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E., Peterson, A.T., 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecol. Model.* 263, 10–18.
- Park, J., Kim, D.I., Choi, B., Kang, W., Kwon, H.W., 2020. Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Sci. Rep.* 10, 1012.
- Parsons, Z., Banitaan, S., 2021. Automatic identification of Chagas disease vectors using data mining and deep learning techniques. *Ecol. Inform.* 62, 101270.
- Pataki, B.A., Garriga, J., Eritja, R., Palmer, J.R.B., Bartumeus, F., Csabai, I., 2021. Deep learning identification for citizen science surveillance of tiger mosquitoes. *Sci. Rep.* 11, 4718.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,

- Duchesnay, E., 2011. *Scikit-learn: Machine learning in Python*. J. Mach. Learn. Res. 12, 2825–2830.
- Pérez-Molina, J.A., Molina, I., 2018. Chagas disease. Lancet 391, 82–94.
- Posit Software. 2022. RStudio 2023.03.1.446. <https://posit.co/products/open-source/rstudio/>.
- Prata, A., 2001. Clinical and epidemiological aspects of Chagas disease. Lancet Infect. Dis. 1, 92–100.
- R Core Team, 2022. R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ribeiro-Jr, G., Abad-Franch, F., de Sousa, O.M., Dos Santos, C.G., Fonseca, E.O., dos Santos, R.F., Cunha, G.M., de Carvalho, C.M.M., Reis, R.B., Gurgel-Gonçalves, R., Reis, M.G., 2021. TriatoScore: An entomological-risk score for Chagas disease vector control-surveillance. Parasit. Vectors, 14, 492.
- Rojas de Arias, A., Monroy, C., Guhl, F., Sosa-Estani, S., Santos, W.S., Abad-Franch, F., 2022. Chagas disease control-surveillance in the Americas: The multinational initiatives and the practical impossibility of interrupting vector-borne *Trypanosoma cruzi* transmission. Mem. Inst. Oswaldo Cruz 117, e210130.
- Schofield, C.J., Jannin, J., Salvatella, R., 2006. The future of Chagas disease control. Trends Parasitol. 22, 583–588.
- Silveira, A.C., Feitosa, V.R., Borges, R., 1984. Distribuição de triatomíneos capturados no ambiente domiciliar, no período 1975/83, Brasil. Rev. Bras. Malariol. Doenças Trop. 36, 15–312.
- Terry, J.C.D., Roy, H.E., August, T.A., 2020. Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. Methods Ecol. Evol. 11, 303–315.

- Tharwat, A., Gaber, T., Ibrahim, A., Hassanien, A.E., 2017. Linear discriminant analysis: A detailed tutorial. *AI Commun.* 30, 169–190.
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecol. Monogr.*, 92, e01486.
- Vinhaes, M.C., de Oliveira, S.V., Reis, P.O., de Lacerda Sousa, A.C., Albuquerque, R., Obara, M.T., Bezerra, C.M., Costa, V.M., Alves, R.V., Gurgel-Gonçalves, R., 2014. Assessing the vulnerability of Brazilian municipalities to the vectorial transmission of *Trypanosoma cruzi* using multi-criteria decision analysis. *Acta Trop.* 137, 105–110.
- WHO – World Health Organization, 2002. Control of Chagas disease: Second report of the WHO Expert Committee. WHO Tech. Rep. Ser. 905, 1–109.
- WHO – World Health Organization, 2017. Global vector control response 2017–2030. World Health Organization, Geneva.
- Yadon, Z., Gürtler, R.E., Tobar, F., Medici, A.C. (Eds.), 2006. Descentralización y Gestión del Control de Enfermedades Transmisibles en América Latina. PAHO/WHO, Buenos Aires.
- Yoshioka, K., Tercero, D., Pérez, B., Nakamura, J., Perez, L., 2017. Implementing a vector surveillance-response system for Chagas disease control: A 4-year field trial in Nicaragua. *Infect. Dis. Poverty*, 6, 18.
- Zhao, G., Pietikainen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 915–928.

Supplementary tables

Tables S1–S15.

Table S1. Data underlying Fig. 2: observed accuracy by algorithm (with score 95% confidence interval limits).

Algorithm	“N”	Correct	Proportion	CI lower	CI upper
AlexNet	13,140	13,140	1.000	0.999	1.000
AdaBoost	13,140	9086	0.691	0.684	0.699
Gradient Boosting	13,140	12,678	0.965	0.962	0.968
Histogram Boosting	13,140	12,825	0.976	0.973	0.979
Linear Discriminant	13,140	12,600	0.959	0.955	0.962

“N”, number of pictures used for testing over 10 (pseudo-)replicate training-testing runs by each machine-learning algorithm

Table S2. Data underlying Fig. 2: observed accuracy by genus (with score 95% confidence interval limits).

Genus	“N”	Correct	Proportion	CI lower	CI upper
<i>Triatoma</i>	35,965	34,791	0.967	0.965	0.969
<i>Panstrongylus</i>	9930	8487	0.855	0.848	0.861
<i>Rhodnius</i>	19,805	17,051	0.861	0.856	0.866

“N”, number of pictures used for testing over 10 (pseudo-)replicate training-testing runs by five machine-learning algorithms

Table S3. Data underlying Fig. 2: observed accuracy by picture angle (with score 95% confidence interval limits).

Angle*	“N”	Correct	Proportion	CI lower	CI upper
0	7190	6652	0.925	0.919	0.931
1	7235	6663	0.921	0.914	0.927
2	7140	6552	0.918	0.911	0.924
3	7430	6885	0.927	0.920	0.932
4	7565	6879	0.909	0.903	0.916
5	7445	6847	0.920	0.913	0.926
6	7390	6781	0.918	0.911	0.924
7	7190	6545	0.910	0.903	0.917
8	7115	6525	0.917	0.910	0.923

*See Fig 1 of main text; “N”, number of pictures used for testing over 10 (pseudo-)replicate training-testing runs by five machine-learning algorithms

Table S4. Data underlying Fig. 2: observed accuracy by position of the bug in the picture (with score 95% confidence interval limits).

Position*	“N”	Correct	Proportion	CI lower	CI upper
A – Dorsal-flat	7190	6652	0.925	0.919	0.931
B – Dorsal-oblique	43,830	40,167	0.916	0.914	0.919
C – Front/back-oblique	14,680	13,510	0.920	0.916	0.925

*See Fig 1 of main text; “N”, number of pictures used for testing over 10 (pseudo-)replicate training-testing runs by five machine-learning algorithms

Table S5. Data underlying Fig. 2: observed accuracy by algorithm and (pseudo-)replicate training-testing run (with score 95% confidence interval limits).

Algorithm	Replicate	“N”	Correct	Proportion	CI lower	CI upper
AlexNet	PSR_1	1314	1314	1.000	0.997	1.000
	PSR_2	1314	1314	1.000	0.997	1.000
	PSR_3	1314	1314	1.000	0.997	1.000
	PSR_4	1314	1314	1.000	0.997	1.000
	PSR_5	1314	1314	1.000	0.997	1.000
	PSR_6	1314	1314	1.000	0.997	1.000
	PSR_7	1314	1314	1.000	0.997	1.000
	PSR_8	1314	1314	1.000	0.997	1.000
	PSR_9	1314	1314	1.000	0.997	1.000
	PSR_10	1314	1314	1.000	0.997	1.000
AdaBoost	PSR_1	1314	957	0.728	0.704	0.752
	PSR_2	1314	840	0.639	0.613	0.665
	PSR_3	1314	925	0.704	0.679	0.728
	PSR_4	1314	955	0.727	0.702	0.750
	PSR_5	1314	887	0.675	0.649	0.700
	PSR_6	1314	954	0.726	0.701	0.749
	PSR_7	1314	999	0.760	0.736	0.783
	PSR_8	1314	792	0.603	0.576	0.629
	PSR_9	1314	782	0.595	0.568	0.621
	PSR_10	1314	995	0.757	0.733	0.780
Gradient Boosting	PSR_1	1314	1270	0.967	0.955	0.975
	PSR_2	1314	1265	0.963	0.951	0.972
	PSR_3	1314	1269	0.966	0.954	0.974
	PSR_4	1314	1269	0.966	0.954	0.974
	PSR_5	1314	1271	0.967	0.956	0.976
	PSR_6	1314	1265	0.963	0.951	0.972
	PSR_7	1314	1268	0.965	0.954	0.974
	PSR_8	1314	1273	0.969	0.958	0.977
	PSR_9	1314	1267	0.964	0.953	0.973
	PSR_10	1314	1261	0.960	0.948	0.969
Histogram Boosting	PSR_1	1314	1282	0.976	0.966	0.983
	PSR_2	1314	1283	0.976	0.967	0.983
	PSR_3	1314	1285	0.978	0.968	0.985
	PSR_4	1314	1281	0.975	0.965	0.982
	PSR_5	1314	1285	0.978	0.968	0.985
	PSR_6	1314	1284	0.977	0.968	0.984
	PSR_7	1314	1282	0.976	0.966	0.983
	PSR_8	1314	1291	0.982	0.974	0.988
	PSR_9	1314	1278	0.973	0.962	0.980
	PSR_10	1314	1274	0.970	0.959	0.978
Linear Discriminant	PSR_1	1314	1254	0.954	0.942	0.964
	PSR_2	1314	1257	0.957	0.944	0.966
	PSR_3	1314	1259	0.958	0.946	0.968
	PSR_4	1314	1258	0.957	0.945	0.967
	PSR_5	1314	1260	0.959	0.947	0.968
	PSR_6	1314	1262	0.960	0.948	0.970
	PSR_7	1314	1263	0.961	0.949	0.970
	PSR_8	1314	1261	0.960	0.948	0.969
	PSR_9	1314	1262	0.960	0.948	0.970
	PSR_10	1314	1264	0.962	0.950	0.971

“N”, number of pictures used for testing in each of the 10 (pseudo-)replicate training-testing runs; Replicate: each of the 10 (pseudo-)replicate training-testing runs each algorithm completed; all 5 algorithms used the same training and testing datasets in each (pseudo-)replicate run

Table S6. Observed accuracy, fully stratified (with score 95% confidence interval limits).

Algorithm	Genus	(Pseudo-)replicate	Position	Angle	Accuracy				
					“N”	Correct	Proportion	CI lower	CI upper
AlexNet (AN)	<i>Triatoma</i>	(Pseudo-)replicate 1	Dorsal-flat	0	84	84	1.000	0.956	1.000
			Sub-total		84	84	1.000	0.956	1.000
			Dorsal-oblique	2	81	81	1.000	0.955	1.000
				3	88	88	1.000	0.958	1.000
				4	92	92	1.000	0.960	1.000
				6	80	80	1.000	0.954	1.000
				7	83	83	1.000	0.956	1.000
				8	78	78	1.000	0.953	1.000
			Sub-total		502	502	1.000	0.992	1.000
			Front/back-oblique	1	68	68	1.000	0.947	1.000
				5	72	72	1.000	0.949	1.000
			Sub-total		140	140	1.000	0.973	1.000
		Sub-total		726	726	1.000	0.995	1.000	
		(Pseudo-)replicate 2	Dorsal-flat	0	72	72	1.000	0.949	1.000
			Sub-total		72	72	1.000	0.949	1.000
			Dorsal-oblique	2	89	89	1.000	0.959	1.000
				3	73	73	1.000	0.950	1.000
				4	92	92	1.000	0.960	1.000
				6	77	77	1.000	0.952	1.000
				7	78	78	1.000	0.953	1.000
				8	88	88	1.000	0.958	1.000
			Sub-total		497	497	1.000	0.992	1.000
			Front/back-oblique	1	83	83	1.000	0.956	1.000
				5	83	83	1.000	0.956	1.000
	Sub-total		166	166	1.000	0.977	1.000		
	Sub-total		735	735	1.000	0.995	1.000		
(Pseudo-)replicate 3	Dorsal-flat	0	72	72	1.000	0.949	1.000		
	Sub-total		72	72	1.000	0.949	1.000		
	Dorsal-oblique	2	75	75	1.000	0.951	1.000		
		3	77	77	1.000	0.952	1.000		
		4	90	90	1.000	0.959	1.000		
		6	83	83	1.000	0.956	1.000		
		7	71	71	1.000	0.949	1.000		
		8	65	65	1.000	0.944	1.000		

			Sub-total		461	461	1.000	0.992	1.000
			Front/back-oblique	1	83	83	1.000	0.956	1.000
				5	93	93	1.000	0.960	1.000
			Sub-total		176	176	1.000	0.979	1.000
		Sub-total			709	709	1.000	0.995	1.000
		(Pseudo-)replicate 4	Dorsal-flat	0	78	78	1.000	0.953	1.000
			Sub-total		78	78	1.000	0.953	1.000
			Dorsal-oblique	2	78	78	1.000	0.953	1.000
				3	87	87	1.000	0.958	1.000
				4	75	75	1.000	0.951	1.000
				6	82	82	1.000	0.955	1.000
				7	77	77	1.000	0.952	1.000
				8	84	84	1.000	0.956	1.000
			Sub-total		483	483	1.000	0.992	1.000
			Front/back-oblique	1	75	75	1.000	0.951	1.000
				5	74	74	1.000	0.951	1.000
			Sub-total		149	149	1.000	0.975	1.000
		Sub-total			710	710	1.000	0.995	1.000
		(Pseudo-)replicate 5	Dorsal-flat	0	73	73	1.000	0.950	1.000
			Sub-total		73	73	1.000	0.950	1.000
			Dorsal-oblique	2	71	71	1.000	0.949	1.000
				3	82	82	1.000	0.955	1.000
				4	88	88	1.000	0.958	1.000
				6	78	78	1.000	0.953	1.000
				7	100	100	1.000	0.963	1.000
				8	76	76	1.000	0.952	1.000
			Sub-total		495	495	1.000	0.992	1.000
			Front/back-oblique	1	72	72	1.000	0.949	1.000
				5	66	66	1.000	0.945	1.000
			Sub-total		138	138	1.000	0.973	1.000
		Sub-total			706	706	1.000	0.995	1.000
		(Pseudo-)replicate 6	Dorsal-flat	0	80	80	1.000	0.954	1.000
			Sub-total		80	80	1.000	0.954	1.000
			Dorsal-oblique	2	76	76	1.000	0.952	1.000
				3	91	91	1.000	0.959	1.000
				4	80	80	1.000	0.954	1.000
				6	80	80	1.000	0.954	1.000
				7	85	85	1.000	0.957	1.000

				8	81	81	1.000	0.955	1.000
			Sub-total		493	493	1.000	0.992	1.000
			Front/back-oblique	1	96	96	1.000	0.962	1.000
				5	89	89	1.000	0.959	1.000
			Sub-total		185	185	1.000	0.980	1.000
			Sub-total		758	758	1.000	0.995	1.000
		(Pseudo-)replicate 7	Dorsal-flat	0	87	87	1.000	0.958	1.000
			Sub-total		87	87	1.000	0.958	1.000
			Dorsal-oblique	2	74	74	1.000	0.951	1.000
				3	87	87	1.000	0.958	1.000
				4	76	76	1.000	0.952	1.000
				6	73	73	1.000	0.950	1.000
				7	77	77	1.000	0.952	1.000
				8	86	86	1.000	0.957	1.000
			Sub-total		473	473	1.000	0.992	1.000
			Front/back-oblique	1	82	82	1.000	0.955	1.000
				5	69	69	1.000	0.947	1.000
			Sub-total		151	151	1.000	0.975	1.000
			Sub-total		711	711	1.000	0.995	1.000
		(Pseudo-)replicate 8	Dorsal-flat	0	91	91	1.000	0.959	1.000
			Sub-total		91	91	1.000	0.959	1.000
			Dorsal-oblique	2	71	71	1.000	0.949	1.000
				3	84	84	1.000	0.956	1.000
				4	66	66	1.000	0.945	1.000
				6	78	78	1.000	0.953	1.000
				7	80	80	1.000	0.954	1.000
				8	62	62	1.000	0.942	1.000
			Sub-total		441	441	1.000	0.991	1.000
			Front/back-oblique	1	77	77	1.000	0.952	1.000
				5	98	98	1.000	0.962	1.000
			Sub-total		175	175	1.000	0.979	1.000
			Sub-total		707	707	1.000	0.995	1.000
		(Pseudo-)replicate 9	Dorsal-flat	0	79	79	1.000	0.954	1.000
			Sub-total		79	79	1.000	0.954	1.000
			Dorsal-oblique	2	90	90	1.000	0.959	1.000
				3	85	85	1.000	0.957	1.000
				4	96	96	1.000	0.962	1.000
				6	84	84	1.000	0.956	1.000

				7	63	63	1.000	0.943	1.000
				8	69	69	1.000	0.947	1.000
			Sub-total		487	487	1.000	0.992	1.000
			Front/back-oblique	1	64	64	1.000	0.943	1.000
				5	72	72	1.000	0.949	1.000
			Sub-total		136	136	1.000	0.973	1.000
		Sub-total			702	702	1.000	0.995	1.000
		(Pseudo-)replicate 10	Dorsal-flat	0	71	71	1.000	0.949	1.000
			Sub-total		71	71	1.000	0.949	1.000
			Dorsal-oblique	2	73	73	1.000	0.950	1.000
				3	81	81	1.000	0.955	1.000
				4	83	83	1.000	0.956	1.000
				6	103	103	1.000	0.964	1.000
				7	63	63	1.000	0.943	1.000
				8	84	84	1.000	0.956	1.000
			Sub-total		487	487	1.000	0.992	1.000
			Front/back-oblique	1	76	76	1.000	0.952	1.000
				5	95	95	1.000	0.961	1.000
			Sub-total		171	171	1.000	0.978	1.000
		Sub-total			729	729	1.000	0.995	1.000
	Sub-total				7193	7193	1.000	0.999	1.000
	<i>Panstrongylus</i>	(Pseudo-)replicate 1	Dorsal-flat	0	20	20	1.000	0.839	1.000
			Sub-total		20	20	1.000	0.839	1.000
			Dorsal-oblique	2	22	22	1.000	0.851	1.000
				3	25	25	1.000	0.867	1.000
				4	20	20	1.000	0.839	1.000
				6	23	23	1.000	0.857	1.000
				7	25	25	1.000	0.867	1.000
				8	29	29	1.000	0.883	1.000
			Sub-total		144	144	1.000	0.974	1.000
			Front/back-oblique	1	20	20	1.000	0.839	1.000
				5	18	18	1.000	0.824	1.000
			Sub-total		38	38	1.000	0.908	1.000
		Sub-total			202	202	1.000	0.981	1.000
		(Pseudo-)replicate 2	Dorsal-flat	0	22	22	1.000	0.851	1.000
			Sub-total		22	22	1.000	0.851	1.000
			Dorsal-oblique	2	20	20	1.000	0.839	1.000
				3	17	17	1.000	0.816	1.000

				4	27	27	1.000	0.875	1.000
				6	23	23	1.000	0.857	1.000
				7	22	22	1.000	0.851	1.000
				8	23	23	1.000	0.857	1.000
			Sub-total		132	132	1.000	0.972	1.000
			Front/back-oblique	1	20	20	1.000	0.839	1.000
				5	25	25	1.000	0.867	1.000
			Sub-total		45	45	1.000	0.921	1.000
		Sub-total			199	199	1.000	0.981	1.000
		(Pseudo-)replicate 3	Dorsal-flat	0	22	22	1.000	0.851	1.000
			Sub-total		22	22	1.000	0.851	1.000
			Dorsal-oblique	2	17	17	1.000	0.816	1.000
				3	17	17	1.000	0.816	1.000
				4	17	17	1.000	0.816	1.000
				6	20	20	1.000	0.839	1.000
				7	24	24	1.000	0.862	1.000
				8	27	27	1.000	0.875	1.000
			Sub-total		122	122	1.000	0.969	1.000
			Front/back-oblique	1	23	23	1.000	0.857	1.000
				5	24	24	1.000	0.862	1.000
			Sub-total		47	47	1.000	0.924	1.000
		Sub-total			191	191	1.000	0.980	1.000
		(Pseudo-)replicate 4	Dorsal-flat	0	23	23	1.000	0.857	1.000
			Sub-total		23	23	1.000	0.857	1.000
			Dorsal-oblique	2	24	24	1.000	0.862	1.000
				3	29	29	1.000	0.883	1.000
				4	24	24	1.000	0.862	1.000
				6	27	27	1.000	0.875	1.000
				7	24	24	1.000	0.862	1.000
				8	22	22	1.000	0.851	1.000
			Sub-total		150	150	1.000	0.975	1.000
			Front/back-oblique	1	34	34	1.000	0.898	1.000
				5	24	24	1.000	0.862	1.000
			Sub-total		58	58	1.000	0.938	1.000
		Sub-total			231	231	1.000	0.984	1.000
		(Pseudo-)replicate 5	Dorsal-flat	0	14	14	1.000	0.785	1.000
			Sub-total		14	14	1.000	0.785	1.000
			Dorsal-oblique	2	28	28	1.000	0.879	1.000

				3	21	21	1.000	0.845	1.000
				4	22	22	1.000	0.851	1.000
				6	27	27	1.000	0.875	1.000
				7	12	12	1.000	0.758	1.000
				8	22	22	1.000	0.851	1.000
			Sub-total		132	132	1.000	0.972	1.000
			Front/back-oblique	1	26	26	1.000	0.871	1.000
				5	24	24	1.000	0.862	1.000
			Sub-total		50	50	1.000	0.929	1.000
		Sub-total			196	196	1.000	0.981	1.000
		(Pseudo-)replicate 6	Dorsal-flat	0	24	24	1.000	0.862	1.000
			Sub-total		24	24	1.000	0.862	1.000
			Dorsal-oblique	2	20	20	1.000	0.839	1.000
				3	18	18	1.000	0.824	1.000
				4	18	18	1.000	0.824	1.000
				6	19	19	1.000	0.832	1.000
				7	26	26	1.000	0.871	1.000
				8	14	14	1.000	0.785	1.000
			Sub-total		115	115	1.000	0.968	1.000
			Front/back-oblique	1	25	25	1.000	0.867	1.000
				5	17	17	1.000	0.816	1.000
			Sub-total		42	42	1.000	0.916	1.000
		Sub-total			181	181	1.000	0.979	1.000
		(Pseudo-)replicate 7	Dorsal-flat	0	23	23	1.000	0.857	1.000
			Sub-total		23	23	1.000	0.857	1.000
			Dorsal-oblique	2	19	19	1.000	0.832	1.000
				3	27	27	1.000	0.875	1.000
				4	19	19	1.000	0.832	1.000
				6	28	28	1.000	0.879	1.000
				7	26	26	1.000	0.871	1.000
				8	14	14	1.000	0.785	1.000
			Sub-total		133	133	1.000	0.972	1.000
			Front/back-oblique	1	23	23	1.000	0.857	1.000
				5	20	20	1.000	0.839	1.000
			Sub-total		43	43	1.000	0.918	1.000
		Sub-total			199	199	1.000	0.981	1.000
		(Pseudo-)replicate 8	Dorsal-flat	0	26	26	1.000	0.871	1.000
			Sub-total		26	26	1.000	0.871	1.000

			Dorsal-oblique	2	21	21	1.000	0.845	1.000
				3	23	23	1.000	0.857	1.000
				4	31	31	1.000	0.890	1.000
				6	19	19	1.000	0.832	1.000
				7	25	25	1.000	0.867	1.000
				8	18	18	1.000	0.824	1.000
			Sub-total		137	137	1.000	0.973	1.000
			Front/back-oblique	1	16	16	1.000	0.806	1.000
				5	28	28	1.000	0.879	1.000
			Sub-total		44	44	1.000	0.920	1.000
		Sub-total			207	207	1.000	0.982	1.000
		(Pseudo-)replicate 9	Dorsal-flat	0	23	23	1.000	0.857	1.000
			Sub-total		23	23	1.000	0.857	1.000
			Dorsal-oblique	2	11	11	1.000	0.741	1.000
				3	23	23	1.000	0.857	1.000
				4	32	32	1.000	0.893	1.000
				6	18	18	1.000	0.824	1.000
				7	18	18	1.000	0.824	1.000
				8	27	27	1.000	0.875	1.000
			Sub-total		129	129	1.000	0.971	1.000
			Front/back-oblique	1	19	19	1.000	0.832	1.000
				5	25	25	1.000	0.867	1.000
			Sub-total		44	44	1.000	0.920	1.000
		Sub-total			196	196	1.000	0.981	1.000
		(Pseudo-)replicate 10	Dorsal-flat	0	21	21	1.000	0.845	1.000
			Sub-total		21	21	1.000	0.845	1.000
			Dorsal-oblique	2	23	23	1.000	0.857	1.000
				3	20	20	1.000	0.839	1.000
				4	25	25	1.000	0.867	1.000
				6	15	15	1.000	0.796	1.000
				7	22	22	1.000	0.851	1.000
				8	22	22	1.000	0.851	1.000
			Sub-total		127	127	1.000	0.971	1.000
			Front/back-oblique	1	17	17	1.000	0.816	1.000
				5	19	19	1.000	0.832	1.000
			Sub-total		36	36	1.000	0.904	1.000
		Sub-total			184	184	1.000	0.980	1.000
	Sub-total				1986	1986	1.000	0.998	1.000

	<i>Rhodnius</i>	(Pseudo-)replicate 1	Dorsal-flat	0	36	36	1.000	0.904	1.000
			Sub-total		36	36	1.000	0.904	1.000
			Dorsal-oblique	2	46	46	1.000	0.923	1.000
				3	38	38	1.000	0.908	1.000
				4	46	46	1.000	0.923	1.000
				6	33	33	1.000	0.896	1.000
				7	51	51	1.000	0.930	1.000
				8	45	45	1.000	0.921	1.000
			Sub-total		259	259	1.000	0.985	1.000
			Front/back-oblique	1	47	47	1.000	0.924	1.000
				5	44	44	1.000	0.920	1.000
			Sub-total		91	91	1.000	0.959	1.000
		Sub-total			386	386	1.000	0.990	1.000
		(Pseudo-)replicate 2	Dorsal-flat	0	46	46	1.000	0.923	1.000
			Sub-total		46	46	1.000	0.923	1.000
			Dorsal-oblique	2	43	43	1.000	0.918	1.000
				3	32	32	1.000	0.893	1.000
				4	42	42	1.000	0.916	1.000
				6	44	44	1.000	0.920	1.000
				7	52	52	1.000	0.931	1.000
				8	40	40	1.000	0.912	1.000
			Sub-total		253	253	1.000	0.985	1.000
			Front/back-oblique	1	47	47	1.000	0.924	1.000
				5	34	34	1.000	0.898	1.000
			Sub-total		81	81	1.000	0.955	1.000
		Sub-total			380	380	1.000	0.990	1.000
		(Pseudo-)replicate 3	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	38	38	1.000	0.908	1.000
				3	50	50	1.000	0.929	1.000
				4	44	44	1.000	0.920	1.000
				6	52	52	1.000	0.931	1.000
				7	50	50	1.000	0.929	1.000
				8	43	43	1.000	0.918	1.000
			Sub-total		277	277	1.000	0.986	1.000
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	52	52	1.000	0.931	1.000
			Sub-total		96	96	1.000	0.962	1.000

		Sub-total			414	414	1.000	0.991	1.000
		(Pseudo-)replicate 4	Dorsal-flat	0	43	43	1.000	0.918	1.000
			Sub-total		43	43	1.000	0.918	1.000
			Dorsal-oblique	2	37	37	1.000	0.906	1.000
				3	51	51	1.000	0.930	1.000
				4	35	35	1.000	0.901	1.000
				6	50	50	1.000	0.929	1.000
				7	32	32	1.000	0.893	1.000
				8	42	42	1.000	0.916	1.000
			Sub-total		247	247	1.000	0.985	1.000
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	39	39	1.000	0.910	1.000
			Sub-total		83	83	1.000	0.956	1.000
		Sub-total			373	373	1.000	0.990	1.000
		(Pseudo-)replicate 5	Dorsal-flat	0	50	50	1.000	0.929	1.000
			Sub-total		50	50	1.000	0.929	1.000
			Dorsal-oblique	2	54	54	1.000	0.934	1.000
				3	40	40	1.000	0.912	1.000
				4	50	50	1.000	0.929	1.000
				6	51	51	1.000	0.930	1.000
				7	49	49	1.000	0.927	1.000
				8	38	38	1.000	0.908	1.000
			Sub-total		282	282	1.000	0.987	1.000
			Front/back-oblique	1	35	35	1.000	0.901	1.000
				5	45	45	1.000	0.921	1.000
			Sub-total		80	80	1.000	0.954	1.000
		Sub-total			412	412	1.000	0.991	1.000
		(Pseudo-)replicate 6	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	43	43	1.000	0.918	1.000
				3	39	39	1.000	0.910	1.000
				4	35	35	1.000	0.901	1.000
				6	40	40	1.000	0.912	1.000
				7	39	39	1.000	0.910	1.000
				8	41	41	1.000	0.914	1.000
			Sub-total		237	237	1.000	0.984	1.000
			Front/back-oblique	1	50	50	1.000	0.929	1.000
				5	47	47	1.000	0.924	1.000

			Sub-total		97	97	1.000	0.962	1.000
		Sub-total			375	375	1.000	0.990	1.000
		(Pseudo-)replicate 7	Dorsal-flat	0	46	46	1.000	0.923	1.000
			Sub-total		46	46	1.000	0.923	1.000
			Dorsal-oblique	2	47	47	1.000	0.924	1.000
				3	45	45	1.000	0.921	1.000
				4	45	45	1.000	0.921	1.000
				6	51	51	1.000	0.930	1.000
				7	34	34	1.000	0.898	1.000
				8	46	46	1.000	0.923	1.000
			Sub-total		268	268	1.000	0.986	1.000
			Front/back-oblique	1	41	41	1.000	0.914	1.000
				5	49	49	1.000	0.927	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total			404	404	1.000	0.991	1.000
		(Pseudo-)replicate 8	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	52	52	1.000	0.931	1.000
				3	43	43	1.000	0.918	1.000
				4	40	40	1.000	0.912	1.000
				6	37	37	1.000	0.906	1.000
				7	46	46	1.000	0.923	1.000
				8	51	51	1.000	0.930	1.000
			Sub-total		269	269	1.000	0.986	1.000
			Front/back-oblique	1	42	42	1.000	0.916	1.000
				5	48	48	1.000	0.926	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total			400	400	1.000	0.990	1.000
		(Pseudo-)replicate 9	Dorsal-flat	0	44	44	1.000	0.920	1.000
			Sub-total		44	44	1.000	0.920	1.000
			Dorsal-oblique	2	41	41	1.000	0.914	1.000
				3	56	56	1.000	0.936	1.000
				4	63	63	1.000	0.943	1.000
				6	38	38	1.000	0.908	1.000
				7	40	40	1.000	0.912	1.000
				8	44	44	1.000	0.920	1.000
			Sub-total		282	282	1.000	0.987	1.000
			Front/back-oblique	1	46	46	1.000	0.923	1.000

				5	44	44	1.000	0.920	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total			416	416	1.000	0.991	1.000
		(Pseudo-)replicate 10	Dorsal-flat	0	45	45	1.000	0.921	1.000
			Sub-total		45	45	1.000	0.921	1.000
			Dorsal-oblique	2	44	44	1.000	0.920	1.000
				3	37	37	1.000	0.906	1.000
				4	40	40	1.000	0.912	1.000
				6	45	45	1.000	0.921	1.000
				7	44	44	1.000	0.920	1.000
				8	42	42	1.000	0.916	1.000
			Sub-total		252	252	1.000	0.985	1.000
			Front/back-oblique	1	52	52	1.000	0.931	1.000
				5	52	52	1.000	0.931	1.000
			Sub-total		104	104	1.000	0.964	1.000
		Sub-total			401	401	1.000	0.991	1.000
	Sub-total				3961	3961	1.000	0.999	1.000
Total AN					13,140	13,140	1.000	1.000	1.000
Multi-class AdaBoost (AB)	<i>Triatoma</i>	(Pseudo-)replicate 1	Dorsal-flat	0	84	83	0.988	0.936	0.998
			Sub-total		84	83	0.988	0.936	0.998
			Dorsal-oblique	2	81	79	0.975	0.914	0.993
				3	88	88	1.000	0.958	1.000
				4	92	89	0.967	0.908	0.989
				6	80	79	0.988	0.933	0.998
				7	83	81	0.976	0.916	0.993
				8	78	77	0.987	0.931	0.998
			Sub-total		502	493	0.982	0.966	0.991
			Front/back-oblique	1	68	67	0.985	0.921	0.997
				5	72	71	0.986	0.925	0.998
			Sub-total		140	138	0.986	0.949	0.996
		Sub-total			726	714	0.983	0.971	0.991
		(Pseudo-)replicate 2	Dorsal-flat	0	72	71	0.986	0.925	0.998
			Sub-total		72	71	0.986	0.925	0.998
			Dorsal-oblique	2	89	84	0.944	0.875	0.976
				3	73	67	0.918	0.832	0.962
				4	92	86	0.935	0.865	0.970
				6	77	72	0.935	0.857	0.972
				7	78	74	0.949	0.875	0.980

			8	88	82	0.932	0.859	0.968	
		Sub-total		497	465	0.936	0.911	0.954	
		Front/back-oblique	1	83	80	0.964	0.899	0.988	
			5	83	79	0.952	0.883	0.981	
		Sub-total		166	159	0.958	0.916	0.979	
		Sub-total		735	695	0.946	0.927	0.960	
		(Pseudo-)replicate 3	Dorsal-flat	0	72	70	0.972	0.904	0.992
			Sub-total		72	70	0.972	0.904	0.992
			Dorsal-oblique	2	75	69	0.920	0.836	0.963
				3	77	71	0.922	0.840	0.964
				4	90	85	0.944	0.876	0.976
				6	83	79	0.952	0.883	0.981
				7	71	63	0.887	0.793	0.942
				8	65	60	0.923	0.832	0.967
			Sub-total		461	427	0.926	0.899	0.947
			Front/back-oblique	1	83	82	0.988	0.935	0.998
				5	93	88	0.946	0.880	0.977
			Sub-total		176	170	0.966	0.928	0.984
		Sub-total		709	667	0.941	0.921	0.956	
		(Pseudo-)replicate 4	Dorsal-flat	0	78	64	0.821	0.721	0.890
			Sub-total		78	64	0.821	0.721	0.890
			Dorsal-oblique	2	78	65	0.833	0.735	0.900
				3	87	67	0.770	0.671	0.846
				4	75	58	0.773	0.667	0.853
				6	82	59	0.720	0.614	0.805
				7	77	62	0.805	0.703	0.878
				8	84	68	0.810	0.713	0.879
			Sub-total		483	379	0.785	0.746	0.819
			Front/back-oblique	1	75	58	0.773	0.667	0.853
				5	74	58	0.784	0.677	0.862
			Sub-total		149	116	0.779	0.705	0.838
		Sub-total		710	559	0.787	0.756	0.816	
		(Pseudo-)replicate 5	Dorsal-flat	0	73	68	0.932	0.849	0.970
			Sub-total		73	68	0.932	0.849	0.970
			Dorsal-oblique	2	71	65	0.915	0.828	0.961
				3	82	80	0.976	0.915	0.993
				4	88	84	0.955	0.889	0.982
				6	78	74	0.949	0.875	0.980

				7	100	94	0.940	0.875	0.972
				8	76	74	0.974	0.909	0.993
			Sub-total		495	471	0.952	0.929	0.967
			Front/back-oblique	1	72	71	0.986	0.925	0.998
				5	66	61	0.924	0.835	0.967
			Sub-total		138	132	0.957	0.908	0.980
		Sub-total			706	671	0.950	0.932	0.964
		(Pseudo-)replicate 6	Dorsal-flat	0	80	79	0.988	0.933	0.998
			Sub-total		80	79	0.988	0.933	0.998
			Dorsal-oblique	2	76	76	1.000	0.952	1.000
				3	91	91	1.000	0.959	1.000
				4	80	76	0.950	0.878	0.980
				6	80	80	1.000	0.954	1.000
				7	85	82	0.965	0.901	0.988
				8	81	80	0.988	0.933	0.998
			Sub-total		493	485	0.984	0.968	0.992
			Front/back-oblique	1	96	96	1.000	0.962	1.000
				5	89	87	0.978	0.922	0.994
			Sub-total		185	183	0.989	0.961	0.997
		Sub-total			758	747	0.985	0.974	0.992
		(Pseudo-)replicate 7	Dorsal-flat	0	87	62	0.713	0.610	0.797
			Sub-total		87	62	0.713	0.610	0.797
			Dorsal-oblique	2	74	64	0.865	0.769	0.925
				3	87	67	0.770	0.671	0.846
				4	76	59	0.776	0.671	0.855
				6	73	57	0.781	0.673	0.860
				7	77	57	0.740	0.633	0.825
				8	86	63	0.733	0.631	0.815
			Sub-total		473	367	0.776	0.736	0.811
			Front/back-oblique	1	82	57	0.695	0.589	0.784
				5	69	56	0.812	0.704	0.886
			Sub-total		151	113	0.748	0.674	0.811
		Sub-total			711	542	0.762	0.730	0.792
		(Pseudo-)replicate 8	Dorsal-flat	0	91	84	0.923	0.850	0.962
			Sub-total		91	84	0.923	0.850	0.962
			Dorsal-oblique	2	71	66	0.930	0.846	0.970
				3	84	82	0.976	0.917	0.993
				4	66	58	0.879	0.779	0.937

				6	78	75	0.962	0.893	0.987
				7	80	73	0.913	0.830	0.957
				8	62	60	0.968	0.890	0.991
			Sub-total		441	414	0.939	0.912	0.958
			Front/back-oblique	1	77	75	0.974	0.910	0.993
				5	98	95	0.969	0.914	0.990
			Sub-total		175	170	0.971	0.935	0.988
		Sub-total			707	668	0.945	0.925	0.959
		(Pseudo-)replicate 9	Dorsal-flat	0	79	79	1.000	0.954	1.000
			Sub-total		79	79	1.000	0.954	1.000
			Dorsal-oblique	2	90	85	0.944	0.876	0.976
				3	85	84	0.988	0.936	0.998
				4	96	90	0.938	0.870	0.971
				6	84	79	0.940	0.868	0.974
				7	63	61	0.968	0.891	0.991
				8	69	65	0.942	0.860	0.977
			Sub-total		487	464	0.953	0.930	0.968
			Front/back-oblique	1	64	60	0.938	0.850	0.975
				5	72	67	0.931	0.848	0.970
			Sub-total		136	127	0.934	0.879	0.965
		Sub-total			702	670	0.954	0.936	0.968
		(Pseudo-)replicate 10	Dorsal-flat	0	71	61	0.859	0.760	0.922
			Sub-total		71	61	0.859	0.760	0.922
			Dorsal-oblique	2	73	58	0.795	0.688	0.871
				3	81	65	0.802	0.703	0.875
				4	83	64	0.771	0.670	0.848
				6	103	79	0.767	0.677	0.838
				7	63	49	0.778	0.661	0.863
				8	84	68	0.810	0.713	0.879
			Sub-total		487	383	0.786	0.748	0.821
			Front/back-oblique	1	76	64	0.842	0.744	0.907
				5	95	81	0.853	0.768	0.910
			Sub-total		171	145	0.848	0.787	0.894
		Sub-total			729	589	0.808	0.778	0.835
	Sub-total				7193	6522	0.907	0.900	0.913
	<i>Panstrongylus</i>	(Pseudo-)replicate 1	Dorsal-flat	0	20	5	0.250	0.112	0.469
			Sub-total		20	5	0.250	0.112	0.469
			Dorsal-oblique	2	22	11	0.500	0.307	0.693

				3	25	7	0.280	0.143	0.476
				4	20	6	0.300	0.145	0.519
				6	23	11	0.478	0.292	0.670
				7	25	10	0.400	0.234	0.593
				8	29	8	0.276	0.147	0.457
			Sub-total		144	53	0.368	0.294	0.449
			Front/back-oblique	1	20	4	0.200	0.081	0.416
				5	18	6	0.333	0.163	0.563
			Sub-total		38	10	0.263	0.150	0.420
		Sub-total			202	68	0.337	0.275	0.404
		(Pseudo-)replicate 2	Dorsal-flat	0	22	19	0.864	0.667	0.953
			Sub-total		22	19	0.864	0.667	0.953
			Dorsal-oblique	2	20	14	0.700	0.481	0.855
				3	17	13	0.765	0.527	0.904
				4	27	12	0.444	0.276	0.627
				6	23	14	0.609	0.408	0.778
				7	22	18	0.818	0.615	0.927
				8	23	13	0.565	0.368	0.744
			Sub-total		132	84	0.636	0.552	0.713
			Front/back-oblique	1	20	14	0.700	0.481	0.855
				5	25	14	0.560	0.371	0.733
			Sub-total		45	28	0.622	0.476	0.749
		Sub-total			199	131	0.658	0.590	0.721
		(Pseudo-)replicate 3	Dorsal-flat	0	22	13	0.591	0.387	0.767
			Sub-total		22	13	0.591	0.387	0.767
			Dorsal-oblique	2	17	9	0.529	0.310	0.738
				3	17	12	0.706	0.469	0.867
				4	17	5	0.294	0.133	0.531
				6	20	13	0.650	0.433	0.819
				7	24	14	0.583	0.388	0.755
				8	27	13	0.481	0.307	0.660
			Sub-total		122	66	0.541	0.453	0.627
			Front/back-oblique	1	23	17	0.739	0.535	0.875
				5	24	9	0.375	0.212	0.573
			Sub-total		47	26	0.553	0.412	0.686
		Sub-total			191	105	0.550	0.479	0.619
		(Pseudo-)replicate 4	Dorsal-flat	0	23	13	0.565	0.368	0.744
			Sub-total		23	13	0.565	0.368	0.744

			Dorsal-oblique	2	24	19	0.792	0.595	0.908
				3	29	22	0.759	0.579	0.878
				4	24	15	0.625	0.427	0.788
				6	27	25	0.926	0.766	0.979
				7	24	21	0.875	0.690	0.957
				8	22	19	0.864	0.667	0.953
			Sub-total		150	121	0.807	0.736	0.862
			Front/back-oblique	1	34	28	0.824	0.665	0.917
				5	24	22	0.917	0.742	0.977
			Sub-total		58	50	0.862	0.751	0.928
		Sub-total			231	184	0.797	0.740	0.843
		(Pseudo-)replicate 5	Dorsal-flat	0	14	9	0.643	0.388	0.837
			Sub-total		14	9	0.643	0.388	0.837
			Dorsal-oblique	2	28	19	0.679	0.493	0.821
				3	21	14	0.667	0.454	0.828
				4	22	4	0.182	0.073	0.385
				6	27	17	0.630	0.442	0.785
				7	12	6	0.500	0.254	0.746
				8	22	11	0.500	0.307	0.693
			Sub-total		132	71	0.538	0.453	0.621
			Front/back-oblique	1	26	19	0.731	0.539	0.863
				5	24	11	0.458	0.279	0.649
			Sub-total		50	30	0.600	0.462	0.724
		Sub-total			196	110	0.561	0.491	0.629
		(Pseudo-)replicate 6	Dorsal-flat	0	24	13	0.542	0.351	0.721
			Sub-total		24	13	0.542	0.351	0.721
			Dorsal-oblique	2	20	12	0.600	0.387	0.781
				3	18	7	0.389	0.203	0.614
				4	18	8	0.444	0.246	0.663
				6	19	9	0.474	0.273	0.683
				7	26	7	0.269	0.137	0.461
				8	14	7	0.500	0.268	0.732
			Sub-total		115	50	0.435	0.348	0.526
			Front/back-oblique	1	25	15	0.600	0.407	0.766
				5	17	4	0.235	0.096	0.473
			Sub-total		42	19	0.452	0.312	0.601
		Sub-total			181	82	0.453	0.382	0.526
		(Pseudo-)replicate 7	Dorsal-flat	0	23	9	0.391	0.222	0.592

			Sub-total		23	9	0.391	0.222	0.592
			Dorsal-oblique	2	19	18	0.947	0.754	0.991
				3	27	17	0.630	0.442	0.785
				4	19	11	0.579	0.363	0.769
				6	28	26	0.929	0.774	0.980
				7	26	22	0.846	0.665	0.938
				8	14	9	0.643	0.388	0.837
			Sub-total		133	103	0.774	0.696	0.837
			Front/back-oblique	1	23	21	0.913	0.732	0.976
				5	20	17	0.850	0.640	0.948
			Sub-total		43	38	0.884	0.755	0.949
		Sub-total			199	150	0.754	0.689	0.808
		(Pseudo-)replicate 8	Dorsal-flat	0	26	17	0.654	0.462	0.806
			Sub-total		26	17	0.654	0.462	0.806
			Dorsal-oblique	2	21	14	0.667	0.454	0.828
				3	23	16	0.696	0.491	0.844
				4	31	15	0.484	0.320	0.652
				6	19	12	0.632	0.410	0.809
				7	25	15	0.600	0.407	0.766
				8	18	9	0.500	0.290	0.710
			Sub-total		137	81	0.591	0.508	0.670
			Front/back-oblique	1	16	10	0.625	0.386	0.815
				5	28	13	0.464	0.295	0.642
			Sub-total		44	23	0.523	0.379	0.662
		Sub-total			207	121	0.585	0.516	0.650
		(Pseudo-)replicate 9	Dorsal-flat	0	23	14	0.609	0.408	0.778
			Sub-total		23	14	0.609	0.408	0.778
			Dorsal-oblique	2	11	5	0.455	0.213	0.720
				3	23	16	0.696	0.491	0.844
				4	32	17	0.531	0.364	0.691
				6	18	9	0.500	0.290	0.710
				7	18	9	0.500	0.290	0.710
				8	27	12	0.444	0.276	0.627
			Sub-total		129	68	0.527	0.441	0.611
			Front/back-oblique	1	19	12	0.632	0.410	0.809
				5	25	13	0.520	0.335	0.700
			Sub-total		44	25	0.568	0.422	0.703
		Sub-total			196	107	0.546	0.476	0.614

		(Pseudo-)replicate 10	Dorsal-flat	0	21	7	0.333	0.172	0.546
			Sub-total		21	7	0.333	0.172	0.546
			Dorsal-oblique	2	23	18	0.783	0.581	0.903
				3	20	11	0.550	0.342	0.742
				4	25	18	0.720	0.524	0.857
				6	15	12	0.800	0.548	0.930
				7	22	14	0.636	0.430	0.803
				8	22	10	0.455	0.269	0.653
			Sub-total		127	83	0.654	0.567	0.731
			Front/back-oblique	1	17	15	0.882	0.657	0.967
				5	19	17	0.895	0.686	0.971
			Sub-total		36	32	0.889	0.747	0.956
		Sub-total			184	122	0.663	0.592	0.727
	Sub-total				1986	1180	0.594	0.572	0.616
	<i>Rhodnius</i>	(Pseudo-)replicate 1	Dorsal-flat	0	36	22	0.611	0.449	0.752
			Sub-total		36	22	0.611	0.449	0.752
			Dorsal-oblique	2	46	17	0.370	0.245	0.514
				3	38	20	0.526	0.373	0.675
				4	46	23	0.500	0.361	0.639
				6	33	14	0.424	0.272	0.592
				7	51	22	0.431	0.305	0.567
				8	45	19	0.422	0.290	0.567
			Sub-total		259	115	0.444	0.385	0.505
			Front/back-oblique	1	47	17	0.362	0.240	0.505
				5	44	21	0.477	0.338	0.621
			Sub-total		91	38	0.418	0.322	0.520
		Sub-total			386	175	0.453	0.404	0.503
		(Pseudo-)replicate 2	Dorsal-flat	0	46	2	0.043	0.012	0.145
			Sub-total		46	2	0.043	0.012	0.145
			Dorsal-oblique	2	43	2	0.047	0.013	0.155
				3	32	2	0.063	0.017	0.201
				4	42	2	0.048	0.013	0.158
				6	44	2	0.045	0.013	0.151
				7	52	1	0.019	0.003	0.101
				8	40	1	0.025	0.004	0.129
			Sub-total		253	10	0.040	0.022	0.071
			Front/back-oblique	1	47	1	0.021	0.004	0.111
				5	34	1	0.029	0.005	0.149

			Sub-total		81	2	0.025	0.007	0.086
		Sub-total			380	14	0.037	0.022	0.061
		(Pseudo-)replicate 3	Dorsal-flat	0	41	28	0.683	0.530	0.804
			Sub-total		41	28	0.683	0.530	0.804
			Dorsal-oblique	2	38	13	0.342	0.212	0.501
				3	50	22	0.440	0.312	0.577
				4	44	14	0.318	0.200	0.466
				6	52	12	0.231	0.137	0.361
				7	50	14	0.280	0.175	0.417
				8	43	20	0.465	0.325	0.611
			Sub-total		277	95	0.343	0.290	0.401
			Front/back-oblique	1	44	11	0.250	0.146	0.394
				5	52	19	0.365	0.248	0.501
			Sub-total		96	30	0.313	0.229	0.411
		Sub-total			414	153	0.370	0.324	0.417
		(Pseudo-)replicate 4	Dorsal-flat	0	43	30	0.698	0.549	0.814
			Sub-total		43	30	0.698	0.549	0.814
			Dorsal-oblique	2	37	18	0.486	0.334	0.641
				3	51	32	0.627	0.490	0.747
				4	35	21	0.600	0.436	0.744
				6	50	23	0.460	0.330	0.596
				7	32	14	0.438	0.282	0.607
				8	42	25	0.595	0.445	0.730
			Sub-total		247	133	0.538	0.476	0.600
			Front/back-oblique	1	44	23	0.523	0.379	0.662
				5	39	26	0.667	0.510	0.794
			Sub-total		83	49	0.590	0.483	0.690
		Sub-total			373	212	0.568	0.518	0.618
		(Pseudo-)replicate 5	Dorsal-flat	0	50	18	0.360	0.241	0.499
			Sub-total		50	18	0.360	0.241	0.499
			Dorsal-oblique	2	54	8	0.148	0.077	0.266
				3	40	14	0.350	0.221	0.505
				4	50	12	0.240	0.143	0.374
				6	51	8	0.157	0.082	0.280
				7	49	12	0.245	0.146	0.381
				8	38	13	0.342	0.212	0.501
			Sub-total		282	67	0.238	0.192	0.291
			Front/back-oblique	1	35	9	0.257	0.142	0.421

			5	45	12	0.267	0.160	0.410	
		Sub-total		80	21	0.263	0.179	0.368	
		Sub-total		412	106	0.257	0.217	0.302	
		(Pseudo-)replicate 6	Dorsal-flat	0	41	27	0.659	0.505	0.784
			Sub-total		41	27	0.659	0.505	0.784
			Dorsal-oblique	2	43	6	0.140	0.066	0.273
				3	39	10	0.256	0.146	0.411
				4	35	12	0.343	0.208	0.508
				6	40	6	0.150	0.071	0.291
				7	39	13	0.333	0.206	0.490
				8	41	12	0.293	0.176	0.445
			Sub-total		237	59	0.249	0.198	0.308
			Front/back-oblique	1	50	14	0.280	0.175	0.417
				5	47	25	0.532	0.392	0.667
			Sub-total		97	39	0.402	0.310	0.502
			Sub-total		375	125	0.333	0.288	0.383
		(Pseudo-)replicate 7	Dorsal-flat	0	46	41	0.891	0.770	0.953
			Sub-total		46	41	0.891	0.770	0.953
			Dorsal-oblique	2	47	38	0.809	0.675	0.896
				3	45	39	0.867	0.738	0.937
				4	45	33	0.733	0.590	0.840
				6	51	34	0.667	0.530	0.780
				7	34	22	0.647	0.479	0.785
				8	46	32	0.696	0.552	0.809
			Sub-total		268	198	0.739	0.683	0.788
			Front/back-oblique	1	41	29	0.707	0.555	0.824
				5	49	39	0.796	0.664	0.885
			Sub-total		90	68	0.756	0.658	0.833
			Sub-total		404	307	0.760	0.716	0.799
		(Pseudo-)replicate 8	Dorsal-flat	0	41	0	0.000	0.000	0.086
			Sub-total		41	0	0.000	0.000	0.086
			Dorsal-oblique	2	52	0	0.000	0.000	0.069
				3	43	1	0.023	0.004	0.121
				4	40	1	0.025	0.004	0.129
				6	37	0	0.000	0.000	0.094
				7	46	0	0.000	0.000	0.077
				8	51	1	0.020	0.003	0.103
			Sub-total		269	3	0.011	0.004	0.032

			Front/back-oblique	1	42	0	0.000	0.000	0.084
				5	48	0	0.000	0.000	0.074
			Sub-total		90	0	0.000	0.000	0.041
		Sub-total			400	3	0.008	0.003	0.022
		(Pseudo-)replicate 9	Dorsal-flat	0	44	1	0.023	0.004	0.118
			Sub-total		44	1	0.023	0.004	0.118
			Dorsal-oblique	2	41	0	0.000	0.000	0.086
				3	56	1	0.018	0.003	0.094
				4	63	2	0.032	0.009	0.109
				6	38	0	0.000	0.000	0.092
				7	40	1	0.025	0.004	0.129
				8	44	0	0.000	0.000	0.080
			Sub-total		282	4	0.014	0.006	0.036
			Front/back-oblique	1	46	0	0.000	0.000	0.077
				5	44	0	0.000	0.000	0.080
			Sub-total		90	0	0.000	0.000	0.041
		Sub-total			416	5	0.012	0.005	0.028
		(Pseudo-)replicate 10	Dorsal-flat	0	45	36	0.800	0.662	0.891
			Sub-total		45	36	0.800	0.662	0.891
			Dorsal-oblique	2	44	34	0.773	0.630	0.872
				3	37	28	0.757	0.599	0.866
				4	40	26	0.650	0.495	0.779
				6	45	26	0.578	0.433	0.710
				7	44	32	0.727	0.582	0.837
				8	42	31	0.738	0.589	0.847
			Sub-total		252	177	0.702	0.643	0.755
			Front/back-oblique	1	52	35	0.673	0.538	0.785
				5	52	36	0.692	0.557	0.801
			Sub-total		104	71	0.683	0.588	0.764
		Sub-total			401	284	0.708	0.662	0.751
	Sub-total				3961	1384	0.349	0.335	0.364
Total AB					13,140	9086	0.691	0.684	0.699
Gradient-based boosting (GB)	<i>Triatoma</i>	(Pseudo-)replicate 1	Dorsal-flat	0	84	82	0.976	0.917	0.993
			Sub-total		84	82	0.976	0.917	0.993
			Dorsal-oblique	2	81	79	0.975	0.914	0.993
				3	88	83	0.943	0.874	0.975
				4	92	91	0.989	0.941	0.998
				6	80	79	0.988	0.933	0.998

				7	83	79	0.952	0.883	0.981
				8	78	77	0.987	0.931	0.998
			Sub-total		502	488	0.972	0.954	0.983
			Front/back-oblique	1	68	66	0.971	0.899	0.992
				5	72	71	0.986	0.925	0.998
			Sub-total		140	137	0.979	0.939	0.993
		Sub-total			726	707	0.974	0.959	0.983
		(Pseudo-)replicate 2	Dorsal-flat	0	72	71	0.986	0.925	0.998
			Sub-total		72	71	0.986	0.925	0.998
			Dorsal-oblique	2	89	82	0.921	0.846	0.961
				3	73	70	0.959	0.886	0.986
				4	92	87	0.946	0.879	0.977
				6	77	76	0.987	0.930	0.998
				7	78	75	0.962	0.893	0.987
				8	88	86	0.977	0.921	0.994
			Sub-total		497	476	0.958	0.936	0.972
			Front/back-oblique	1	83	83	1.000	0.956	1.000
				5	83	81	0.976	0.916	0.993
			Sub-total		166	164	0.988	0.957	0.997
		Sub-total			735	711	0.967	0.952	0.978
		(Pseudo-)replicate 3	Dorsal-flat	0	72	70	0.972	0.904	0.992
			Sub-total		72	70	0.972	0.904	0.992
			Dorsal-oblique	2	75	74	0.987	0.928	0.998
				3	77	73	0.948	0.874	0.980
				4	90	89	0.989	0.940	0.998
				6	83	81	0.976	0.916	0.993
				7	71	67	0.944	0.864	0.978
				8	65	63	0.969	0.895	0.992
			Sub-total		461	447	0.970	0.950	0.982
			Front/back-oblique	1	83	82	0.988	0.935	0.998
				5	93	91	0.978	0.925	0.994
			Sub-total		176	173	0.983	0.951	0.994
		Sub-total			709	690	0.973	0.959	0.983
		(Pseudo-)replicate 4	Dorsal-flat	0	78	74	0.949	0.875	0.980
			Sub-total		78	74	0.949	0.875	0.980
			Dorsal-oblique	2	78	75	0.962	0.893	0.987
				3	87	84	0.966	0.903	0.988
				4	75	73	0.973	0.908	0.993

				6	82	82	1.000	0.955	1.000
				7	77	75	0.974	0.910	0.993
				8	84	84	1.000	0.956	1.000
			Sub-total		483	473	0.979	0.962	0.989
			Front/back-oblique	1	75	74	0.987	0.928	0.998
				5	74	72	0.973	0.907	0.993
			Sub-total		149	146	0.980	0.942	0.993
		Sub-total			710	693	0.976	0.962	0.985
		(Pseudo-)replicate 5	Dorsal-flat	0	73	69	0.945	0.867	0.978
			Sub-total		73	69	0.945	0.867	0.978
			Dorsal-oblique	2	71	70	0.986	0.924	0.998
				3	82	79	0.963	0.898	0.987
				4	88	86	0.977	0.921	0.994
				6	78	76	0.974	0.911	0.993
				7	100	97	0.970	0.915	0.990
				8	76	74	0.974	0.909	0.993
			Sub-total		495	482	0.974	0.956	0.985
			Front/back-oblique	1	72	71	0.986	0.925	0.998
				5	66	64	0.970	0.896	0.992
			Sub-total		138	135	0.978	0.938	0.993
		Sub-total			706	686	0.972	0.957	0.982
		(Pseudo-)replicate 6	Dorsal-flat	0	80	78	0.975	0.913	0.993
			Sub-total		80	78	0.975	0.913	0.993
			Dorsal-oblique	2	76	73	0.961	0.890	0.986
				3	91	90	0.989	0.940	0.998
				4	80	78	0.975	0.913	0.993
				6	80	79	0.988	0.933	0.998
				7	85	82	0.965	0.901	0.988
				8	81	80	0.988	0.933	0.998
			Sub-total		493	482	0.978	0.960	0.987
			Front/back-oblique	1	96	96	1.000	0.962	1.000
				5	89	85	0.955	0.890	0.982
			Sub-total		185	181	0.978	0.946	0.992
		Sub-total			758	741	0.978	0.964	0.986
		(Pseudo-)replicate 7	Dorsal-flat	0	87	86	0.989	0.938	0.998
			Sub-total		87	86	0.989	0.938	0.998
			Dorsal-oblique	2	74	73	0.986	0.927	0.998
				3	87	84	0.966	0.903	0.988

				4	76	74	0.974	0.909	0.993
				6	73	72	0.986	0.926	0.998
				7	77	76	0.987	0.930	0.998
				8	86	85	0.988	0.937	0.998
			Sub-total		473	464	0.981	0.964	0.990
			Front/back-oblique	1	82	79	0.963	0.898	0.987
				5	69	67	0.971	0.900	0.992
			Sub-total		151	146	0.967	0.925	0.986
		Sub-total			711	696	0.979	0.965	0.987
		(Pseudo-)replicate 8	Dorsal-flat	0	91	90	0.989	0.940	0.998
			Sub-total		91	90	0.989	0.940	0.998
			Dorsal-oblique	2	71	71	1.000	0.949	1.000
				3	84	81	0.964	0.900	0.988
				4	66	65	0.985	0.919	0.997
				6	78	75	0.962	0.893	0.987
				7	80	74	0.925	0.846	0.965
				8	62	62	1.000	0.942	1.000
			Sub-total		441	428	0.971	0.950	0.983
			Front/back-oblique	1	77	75	0.974	0.910	0.993
				5	98	98	1.000	0.962	1.000
			Sub-total		175	173	0.989	0.959	0.997
		Sub-total			707	691	0.977	0.964	0.986
		(Pseudo-)replicate 9	Dorsal-flat	0	79	78	0.987	0.932	0.998
			Sub-total		79	78	0.987	0.932	0.998
			Dorsal-oblique	2	90	88	0.978	0.923	0.994
				3	85	82	0.965	0.901	0.988
				4	96	92	0.958	0.898	0.984
				6	84	82	0.976	0.917	0.993
				7	63	61	0.968	0.891	0.991
				8	69	69	1.000	0.947	1.000
			Sub-total		487	474	0.973	0.955	0.984
			Front/back-oblique	1	64	62	0.969	0.893	0.991
				5	72	70	0.972	0.904	0.992
			Sub-total		136	132	0.971	0.927	0.989
		Sub-total			702	684	0.974	0.960	0.984
		(Pseudo-)replicate 10	Dorsal-flat	0	71	70	0.986	0.924	0.998
			Sub-total		71	70	0.986	0.924	0.998
			Dorsal-oblique	2	73	72	0.986	0.926	0.998

				3	81	77	0.951	0.880	0.981
				4	83	78	0.940	0.867	0.974
				6	103	101	0.981	0.932	0.995
				7	63	60	0.952	0.869	0.984
				8	84	82	0.976	0.917	0.993
			Sub-total		487	470	0.965	0.945	0.978
			Front/back-oblique	1	76	76	1.000	0.952	1.000
				5	95	93	0.979	0.926	0.994
			Sub-total		171	169	0.988	0.958	0.997
		Sub-total			729	709	0.973	0.958	0.982
	Sub-total				7193	7008	0.974	0.970	0.978
	<i>Panstrongylus</i>	(Pseudo-)replicate 1	Dorsal-flat	0	20	19	0.950	0.764	0.991
			Sub-total		20	19	0.950	0.764	0.991
			Dorsal-oblique	2	22	20	0.909	0.722	0.975
				3	25	24	0.960	0.805	0.993
				4	20	19	0.950	0.764	0.991
				6	23	21	0.913	0.732	0.976
				7	25	23	0.920	0.750	0.978
				8	29	29	1.000	0.883	1.000
			Sub-total		144	136	0.944	0.894	0.972
			Front/back-oblique	1	20	15	0.750	0.531	0.888
				5	18	14	0.778	0.548	0.910
			Sub-total		38	29	0.763	0.608	0.870
		Sub-total			202	184	0.911	0.864	0.943
		(Pseudo-)replicate 2	Dorsal-flat	0	22	21	0.955	0.782	0.992
			Sub-total		22	21	0.955	0.782	0.992
			Dorsal-oblique	2	20	18	0.900	0.699	0.972
				3	17	17	1.000	0.816	1.000
				4	27	21	0.778	0.592	0.894
				6	23	20	0.870	0.679	0.955
				7	22	21	0.955	0.782	0.992
				8	23	20	0.870	0.679	0.955
			Sub-total		132	117	0.886	0.821	0.930
			Front/back-oblique	1	20	18	0.900	0.699	0.972
				5	25	23	0.920	0.750	0.978
			Sub-total		45	41	0.911	0.793	0.965
		Sub-total			199	179	0.899	0.850	0.934
		(Pseudo-)replicate 3	Dorsal-flat	0	22	22	1.000	0.851	1.000

			Sub-total		22	22	1.000	0.851	1.000
			Dorsal-oblique	2	17	16	0.941	0.730	0.990
				3	17	15	0.882	0.657	0.967
				4	17	14	0.824	0.590	0.938
				6	20	17	0.850	0.640	0.948
				7	24	21	0.875	0.690	0.957
				8	27	21	0.778	0.592	0.894
			Sub-total		122	104	0.852	0.779	0.905
			Front/back-oblique	1	23	23	1.000	0.857	1.000
				5	24	20	0.833	0.641	0.933
			Sub-total		47	43	0.915	0.801	0.966
		Sub-total			191	169	0.885	0.832	0.923
		(Pseudo-)replicate 4	Dorsal-flat	0	23	20	0.870	0.679	0.955
			Sub-total		23	20	0.870	0.679	0.955
			Dorsal-oblique	2	24	20	0.833	0.641	0.933
				3	29	26	0.897	0.736	0.964
				4	24	21	0.875	0.690	0.957
				6	27	25	0.926	0.766	0.979
				7	24	23	0.958	0.798	0.993
				8	22	21	0.955	0.782	0.992
			Sub-total		150	136	0.907	0.849	0.944
			Front/back-oblique	1	34	29	0.853	0.699	0.936
				5	24	22	0.917	0.742	0.977
			Sub-total		58	51	0.879	0.771	0.940
		Sub-total			231	207	0.896	0.850	0.929
		(Pseudo-)replicate 5	Dorsal-flat	0	14	12	0.857	0.601	0.960
			Sub-total		14	12	0.857	0.601	0.960
			Dorsal-oblique	2	28	25	0.893	0.728	0.963
				3	21	21	1.000	0.845	1.000
				4	22	20	0.909	0.722	0.975
				6	27	25	0.926	0.766	0.979
				7	12	9	0.750	0.468	0.911
				8	22	18	0.818	0.615	0.927
			Sub-total		132	118	0.894	0.830	0.936
			Front/back-oblique	1	26	24	0.923	0.759	0.979
				5	24	24	1.000	0.862	1.000
			Sub-total		50	48	0.960	0.865	0.989
		Sub-total			196	178	0.908	0.860	0.941

		(Pseudo-)replicate 6	Dorsal-flat	0	24	20	0.833	0.641	0.933
			Sub-total		24	20	0.833	0.641	0.933
			Dorsal-oblique	2	20	18	0.900	0.699	0.972
				3	18	18	1.000	0.824	1.000
				4	18	16	0.889	0.672	0.969
				6	19	17	0.895	0.686	0.971
				7	26	22	0.846	0.665	0.938
				8	14	12	0.857	0.601	0.960
			Sub-total		115	103	0.896	0.826	0.939
			Front/back-oblique	1	25	22	0.880	0.700	0.958
				5	17	13	0.765	0.527	0.904
			Sub-total		42	35	0.833	0.694	0.917
			Sub-total		181	158	0.873	0.817	0.914
		(Pseudo-)replicate 7	Dorsal-flat	0	23	21	0.913	0.732	0.976
			Sub-total		23	21	0.913	0.732	0.976
			Dorsal-oblique	2	19	18	0.947	0.754	0.991
				3	27	26	0.963	0.817	0.993
				4	19	17	0.895	0.686	0.971
				6	28	24	0.857	0.685	0.943
				7	26	24	0.923	0.759	0.979
				8	14	11	0.786	0.524	0.924
			Sub-total		133	120	0.902	0.840	0.942
			Front/back-oblique	1	23	19	0.826	0.629	0.930
				5	20	17	0.850	0.640	0.948
			Sub-total		43	36	0.837	0.700	0.919
			Sub-total		199	177	0.889	0.838	0.926
		(Pseudo-)replicate 8	Dorsal-flat	0	26	23	0.885	0.710	0.960
			Sub-total		26	23	0.885	0.710	0.960
			Dorsal-oblique	2	21	18	0.857	0.654	0.950
				3	23	22	0.957	0.790	0.992
				4	31	27	0.871	0.711	0.949
				6	19	17	0.895	0.686	0.971
				7	25	23	0.920	0.750	0.978
				8	18	17	0.944	0.742	0.990
			Sub-total		137	124	0.905	0.844	0.944
			Front/back-oblique	1	16	14	0.875	0.640	0.965
				5	28	24	0.857	0.685	0.943
			Sub-total		44	38	0.864	0.733	0.936

		Sub-total			207	185	0.894	0.844	0.929
		(Pseudo-)replicate 9	Dorsal-flat	0	23	20	0.870	0.679	0.955
			Sub-total		23	20	0.870	0.679	0.955
			Dorsal-oblique	2	11	9	0.818	0.523	0.949
				3	23	22	0.957	0.790	0.992
				4	32	31	0.969	0.843	0.994
				6	18	16	0.889	0.672	0.969
				7	18	17	0.944	0.742	0.990
				8	27	23	0.852	0.675	0.941
			Sub-total		129	118	0.915	0.854	0.952
			Front/back-oblique	1	19	16	0.842	0.624	0.945
				5	25	20	0.800	0.609	0.911
			Sub-total		44	36	0.818	0.680	0.905
		Sub-total			196	174	0.888	0.836	0.925
		(Pseudo-)replicate 10	Dorsal-flat	0	21	19	0.905	0.711	0.973
			Sub-total		21	19	0.905	0.711	0.973
			Dorsal-oblique	2	23	20	0.870	0.679	0.955
				3	20	19	0.950	0.764	0.991
				4	25	21	0.840	0.653	0.936
				6	15	13	0.867	0.621	0.963
				7	22	16	0.727	0.518	0.868
				8	22	19	0.864	0.667	0.953
			Sub-total		127	108	0.850	0.778	0.902
			Front/back-oblique	1	17	15	0.882	0.657	0.967
				5	19	17	0.895	0.686	0.971
			Sub-total		36	32	0.889	0.747	0.956
		Sub-total			184	159	0.864	0.807	0.906
	Sub-total				1986	1770	0.891	0.877	0.904
	<i>Rhodnius</i>	(Pseudo-)replicate 1	Dorsal-flat	0	36	35	0.972	0.858	0.995
			Sub-total		36	35	0.972	0.858	0.995
			Dorsal-oblique	2	46	46	1.000	0.923	1.000
				3	38	37	0.974	0.865	0.995
				4	46	45	0.978	0.887	0.996
				6	33	33	1.000	0.896	1.000
				7	51	50	0.980	0.897	0.997
				8	45	45	1.000	0.921	1.000
			Sub-total		259	256	0.988	0.967	0.996
			Front/back-oblique	1	47	45	0.957	0.858	0.988

			5	44	43	0.977	0.882	0.996	
		Sub-total		91	88	0.967	0.908	0.989	
		Sub-total		386	379	0.982	0.963	0.991	
		(Pseudo-)replicate 2	Dorsal-flat	0	46	44	0.957	0.855	0.988
			Sub-total		46	44	0.957	0.855	0.988
			Dorsal-oblique	2	43	42	0.977	0.879	0.996
				3	32	32	1.000	0.893	1.000
				4	42	42	1.000	0.916	1.000
				6	44	44	1.000	0.920	1.000
				7	52	51	0.981	0.899	0.997
				8	40	40	1.000	0.912	1.000
			Sub-total		253	251	0.992	0.972	0.998
			Front/back-oblique	1	47	46	0.979	0.889	0.996
				5	34	34	1.000	0.898	1.000
			Sub-total		81	80	0.988	0.933	0.998
		Sub-total		380	375	0.987	0.970	0.994	
		(Pseudo-)replicate 3	Dorsal-flat	0	41	40	0.976	0.874	0.996
			Sub-total		41	40	0.976	0.874	0.996
			Dorsal-oblique	2	38	38	1.000	0.908	1.000
				3	50	50	1.000	0.929	1.000
				4	44	44	1.000	0.920	1.000
				6	52	52	1.000	0.931	1.000
				7	50	49	0.980	0.895	0.996
				8	43	43	1.000	0.918	1.000
			Sub-total		277	276	0.996	0.980	0.999
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	52	50	0.962	0.870	0.989
			Sub-total		96	94	0.979	0.927	0.994
		Sub-total		414	410	0.990	0.975	0.996	
		(Pseudo-)replicate 4	Dorsal-flat	0	43	42	0.977	0.879	0.996
			Sub-total		43	42	0.977	0.879	0.996
			Dorsal-oblique	2	37	37	1.000	0.906	1.000
				3	51	51	1.000	0.930	1.000
				4	35	35	1.000	0.901	1.000
				6	50	49	0.980	0.895	0.996
				7	32	32	1.000	0.893	1.000
				8	42	42	1.000	0.916	1.000
			Sub-total		247	246	0.996	0.977	0.999

			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	39	37	0.949	0.831	0.986
			Sub-total		83	81	0.976	0.916	0.993
		Sub-total			373	369	0.989	0.973	0.996
		(Pseudo-)replicate 5	Dorsal-flat	0	50	49	0.980	0.895	0.996
			Sub-total		50	49	0.980	0.895	0.996
			Dorsal-oblique	2	54	52	0.963	0.875	0.990
				3	40	40	1.000	0.912	1.000
				4	50	49	0.980	0.895	0.996
				6	51	50	0.980	0.897	0.997
				7	49	49	1.000	0.927	1.000
				8	38	38	1.000	0.908	1.000
			Sub-total		282	278	0.986	0.964	0.994
			Front/back-oblique	1	35	35	1.000	0.901	1.000
				5	45	45	1.000	0.921	1.000
			Sub-total		80	80	1.000	0.954	1.000
		Sub-total			412	407	0.988	0.972	0.995
		(Pseudo-)replicate 6	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	43	41	0.953	0.845	0.987
				3	39	38	0.974	0.868	0.995
				4	35	34	0.971	0.855	0.995
				6	40	38	0.950	0.835	0.986
				7	39	39	1.000	0.910	1.000
				8	41	40	0.976	0.874	0.996
			Sub-total		237	230	0.970	0.940	0.986
			Front/back-oblique	1	50	49	0.980	0.895	0.996
				5	47	46	0.979	0.889	0.996
			Sub-total		97	95	0.979	0.928	0.994
		Sub-total			375	366	0.976	0.955	0.987
		(Pseudo-)replicate 7	Dorsal-flat	0	46	45	0.978	0.887	0.996
			Sub-total		46	45	0.978	0.887	0.996
			Dorsal-oblique	2	47	45	0.957	0.858	0.988
				3	45	45	1.000	0.921	1.000
				4	45	44	0.978	0.884	0.996
				6	51	48	0.941	0.841	0.980
				7	34	33	0.971	0.851	0.995
				8	46	46	1.000	0.923	1.000

			Sub-total		268	261	0.974	0.947	0.987
			Front/back-oblique	1	41	40	0.976	0.874	0.996
				5	49	49	1.000	0.927	1.000
			Sub-total		90	89	0.989	0.940	0.998
		Sub-total			404	395	0.978	0.958	0.988
		(Pseudo-)replicate 8	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	52	52	1.000	0.931	1.000
				3	43	42	0.977	0.879	0.996
				4	40	40	1.000	0.912	1.000
				6	37	37	1.000	0.906	1.000
				7	46	45	0.978	0.887	0.996
				8	51	51	1.000	0.930	1.000
			Sub-total		269	267	0.993	0.973	0.998
			Front/back-oblique	1	42	41	0.976	0.877	0.996
				5	48	48	1.000	0.926	1.000
			Sub-total		90	89	0.989	0.940	0.998
		Sub-total			400	397	0.993	0.978	0.997
		(Pseudo-)replicate 9	Dorsal-flat	0	44	43	0.977	0.882	0.996
			Sub-total		44	43	0.977	0.882	0.996
			Dorsal-oblique	2	41	40	0.976	0.874	0.996
				3	56	56	1.000	0.936	1.000
				4	63	60	0.952	0.869	0.984
				6	38	38	1.000	0.908	1.000
				7	40	40	1.000	0.912	1.000
				8	44	44	1.000	0.920	1.000
			Sub-total		282	278	0.986	0.964	0.994
			Front/back-oblique	1	46	45	0.978	0.887	0.996
				5	44	43	0.977	0.882	0.996
			Sub-total		90	88	0.978	0.923	0.994
		Sub-total			416	409	0.983	0.966	0.992
		(Pseudo-)replicate 10	Dorsal-flat	0	45	44	0.978	0.884	0.996
			Sub-total		45	44	0.978	0.884	0.996
			Dorsal-oblique	2	44	43	0.977	0.882	0.996
				3	37	37	1.000	0.906	1.000
				4	40	39	0.975	0.871	0.996
				6	45	44	0.978	0.884	0.996
				7	44	43	0.977	0.882	0.996

				8	42	42	1.000	0.916	1.000
			Sub-total		252	248	0.984	0.960	0.994
			Front/back-oblique	1	52	50	0.962	0.870	0.989
				5	52	51	0.981	0.899	0.997
			Sub-total		104	101	0.971	0.919	0.990
		Sub-total			401	393	0.980	0.961	0.990
	Sub-total				3961	3900	0.985	0.980	0.988
Total GB					13,140	12,678	0.965	0.962	0.968
Histogram-based boosting (HB)	<i>Triatoma</i>	(Pseudo-)replicate 1	Dorsal-flat	0	84	83	0.988	0.936	0.998
			Sub-total		84	83	0.988	0.936	0.998
			Dorsal-oblique	2	81	81	1.000	0.955	1.000
				3	88	83	0.943	0.874	0.975
				4	92	91	0.989	0.941	0.998
				6	80	79	0.988	0.933	0.998
				7	83	82	0.988	0.935	0.998
				8	78	78	1.000	0.953	1.000
			Sub-total		502	494	0.984	0.969	0.992
			Front/back-oblique	1	68	66	0.971	0.899	0.992
				5	72	71	0.986	0.925	0.998
			Sub-total		140	137	0.979	0.939	0.993
		Sub-total			726	714	0.983	0.971	0.991
		(Pseudo-)replicate 2	Dorsal-flat	0	72	72	1.000	0.949	1.000
			Sub-total		72	72	1.000	0.949	1.000
			Dorsal-oblique	2	89	86	0.966	0.906	0.988
				3	73	71	0.973	0.905	0.992
				4	92	89	0.967	0.908	0.989
				6	77	76	0.987	0.930	0.998
				7	78	75	0.962	0.893	0.987
				8	88	88	1.000	0.958	1.000
			Sub-total		497	485	0.976	0.958	0.986
			Front/back-oblique	1	83	83	1.000	0.956	1.000
				5	83	82	0.988	0.935	0.998
			Sub-total		166	165	0.994	0.967	0.999
		Sub-total			735	722	0.982	0.970	0.990
		(Pseudo-)replicate 3	Dorsal-flat	0	72	72	1.000	0.949	1.000
			Sub-total		72	72	1.000	0.949	1.000
			Dorsal-oblique	2	75	75	1.000	0.951	1.000
				3	77	74	0.961	0.892	0.987

				4	90	90	1.000	0.959	1.000
				6	83	82	0.988	0.935	0.998
				7	71	70	0.986	0.924	0.998
				8	65	63	0.969	0.895	0.992
			Sub-total		461	454	0.985	0.969	0.993
			Front/back-oblique	1	83	83	1.000	0.956	1.000
				5	93	91	0.978	0.925	0.994
			Sub-total		176	174	0.989	0.960	0.997
		Sub-total			709	700	0.987	0.976	0.993
		(Pseudo-)replicate 4	Dorsal-flat	0	78	75	0.962	0.893	0.987
			Sub-total		78	75	0.962	0.893	0.987
			Dorsal-oblique	2	78	76	0.974	0.911	0.993
				3	87	85	0.977	0.920	0.994
				4	75	74	0.987	0.928	0.998
				6	82	82	1.000	0.955	1.000
				7	77	76	0.987	0.930	0.998
				8	84	84	1.000	0.956	1.000
			Sub-total		483	477	0.988	0.973	0.994
			Front/back-oblique	1	75	74	0.987	0.928	0.998
				5	74	73	0.986	0.927	0.998
			Sub-total		149	147	0.987	0.952	0.996
		Sub-total			710	699	0.985	0.972	0.991
		(Pseudo-)replicate 5	Dorsal-flat	0	73	72	0.986	0.926	0.998
			Sub-total		73	72	0.986	0.926	0.998
			Dorsal-oblique	2	71	70	0.986	0.924	0.998
				3	82	79	0.963	0.898	0.987
				4	88	86	0.977	0.921	0.994
				6	78	76	0.974	0.911	0.993
				7	100	98	0.980	0.930	0.994
				8	76	76	1.000	0.952	1.000
			Sub-total		495	485	0.980	0.963	0.989
			Front/back-oblique	1	72	72	1.000	0.949	1.000
				5	66	64	0.970	0.896	0.992
			Sub-total		138	136	0.986	0.949	0.996
		Sub-total			706	693	0.982	0.969	0.989
		(Pseudo-)replicate 6	Dorsal-flat	0	80	79	0.988	0.933	0.998
			Sub-total		80	79	0.988	0.933	0.998
			Dorsal-oblique	2	76	75	0.987	0.929	0.998

				3	91	89	0.978	0.923	0.994
				4	80	79	0.988	0.933	0.998
				6	80	79	0.988	0.933	0.998
				7	85	84	0.988	0.936	0.998
				8	81	81	1.000	0.955	1.000
			Sub-total		493	487	0.988	0.974	0.994
			Front/back-oblique	1	96	96	1.000	0.962	1.000
				5	89	87	0.978	0.922	0.994
			Sub-total		185	183	0.989	0.961	0.997
		Sub-total			758	749	0.988	0.978	0.994
		(Pseudo-)replicate 7	Dorsal-flat	0	87	87	1.000	0.958	1.000
			Sub-total		87	87	1.000	0.958	1.000
			Dorsal-oblique	2	74	74	1.000	0.951	1.000
				3	87	84	0.966	0.903	0.988
				4	76	76	1.000	0.952	1.000
				6	73	73	1.000	0.950	1.000
				7	77	77	1.000	0.952	1.000
				8	86	86	1.000	0.957	1.000
			Sub-total		473	470	0.994	0.982	0.998
			Front/back-oblique	1	82	82	1.000	0.955	1.000
				5	69	68	0.986	0.922	0.997
			Sub-total		151	150	0.993	0.963	0.999
		Sub-total			711	707	0.994	0.986	0.998
		(Pseudo-)replicate 8	Dorsal-flat	0	91	90	0.989	0.940	0.998
			Sub-total		91	90	0.989	0.940	0.998
			Dorsal-oblique	2	71	71	1.000	0.949	1.000
				3	84	82	0.976	0.917	0.993
				4	66	66	1.000	0.945	1.000
				6	78	75	0.962	0.893	0.987
				7	80	77	0.963	0.895	0.987
				8	62	62	1.000	0.942	1.000
			Sub-total		441	433	0.982	0.965	0.991
			Front/back-oblique	1	77	76	0.987	0.930	0.998
				5	98	97	0.990	0.944	0.998
			Sub-total		175	173	0.989	0.959	0.997
		Sub-total			707	696	0.984	0.972	0.991
		(Pseudo-)replicate 9	Dorsal-flat	0	79	79	1.000	0.954	1.000
			Sub-total		79	79	1.000	0.954	1.000

			Dorsal-oblique	2	90	87	0.967	0.907	0.989
				3	85	85	1.000	0.957	1.000
				4	96	91	0.948	0.884	0.978
				6	84	84	1.000	0.956	1.000
				7	63	62	0.984	0.915	0.997
				8	69	68	0.986	0.922	0.997
			Sub-total		487	477	0.979	0.963	0.989
			Front/back-oblique	1	64	64	1.000	0.943	1.000
				5	72	70	0.972	0.904	0.992
			Sub-total		136	134	0.985	0.948	0.996
		Sub-total			702	690	0.983	0.970	0.990
		(Pseudo-)replicate 10	Dorsal-flat	0	71	71	1.000	0.949	1.000
			Sub-total		71	71	1.000	0.949	1.000
			Dorsal-oblique	2	73	72	0.986	0.926	0.998
				3	81	80	0.988	0.933	0.998
				4	83	79	0.952	0.883	0.981
				6	103	100	0.971	0.918	0.990
				7	63	62	0.984	0.915	0.997
				8	84	83	0.988	0.936	0.998
			Sub-total		487	476	0.977	0.960	0.987
			Front/back-oblique	1	76	76	1.000	0.952	1.000
				5	95	93	0.979	0.926	0.994
			Sub-total		171	169	0.988	0.958	0.997
		Sub-total			729	716	0.982	0.970	0.990
		Sub-total			7193	7086	0.985	0.982	0.988
	<i>Panstrongylus</i>	(Pseudo-)replicate 1	Dorsal-flat	0	20	19	0.950	0.764	0.991
			Sub-total		20	19	0.950	0.764	0.991
			Dorsal-oblique	2	22	20	0.909	0.722	0.975
				3	25	24	0.960	0.805	0.993
				4	20	17	0.850	0.640	0.948
				6	23	22	0.957	0.790	0.992
				7	25	24	0.960	0.805	0.993
				8	29	29	1.000	0.883	1.000
			Sub-total		144	136	0.944	0.894	0.972
			Front/back-oblique	1	20	16	0.800	0.584	0.919
				5	18	16	0.889	0.672	0.969
			Sub-total		38	32	0.842	0.696	0.926
		Sub-total			202	187	0.926	0.881	0.954

		(Pseudo-)replicate 2	Dorsal-flat	0	22	21	0.955	0.782	0.992
			Sub-total		22	21	0.955	0.782	0.992
			Dorsal-oblique	2	20	17	0.850	0.640	0.948
				3	17	17	1.000	0.816	1.000
				4	27	23	0.852	0.675	0.941
				6	23	21	0.913	0.732	0.976
				7	22	21	0.955	0.782	0.992
				8	23	22	0.957	0.790	0.992
			Sub-total		132	121	0.917	0.857	0.953
			Front/back-oblique	1	20	18	0.900	0.699	0.972
				5	25	24	0.960	0.805	0.993
			Sub-total		45	42	0.933	0.821	0.977
		Sub-total			199	184	0.925	0.879	0.954
		(Pseudo-)replicate 3	Dorsal-flat	0	22	22	1.000	0.851	1.000
			Sub-total		22	22	1.000	0.851	1.000
			Dorsal-oblique	2	17	16	0.941	0.730	0.990
				3	17	16	0.941	0.730	0.990
				4	17	14	0.824	0.590	0.938
				6	20	18	0.900	0.699	0.972
				7	24	21	0.875	0.690	0.957
				8	27	23	0.852	0.675	0.941
			Sub-total		122	108	0.885	0.817	0.930
			Front/back-oblique	1	23	23	1.000	0.857	1.000
				5	24	20	0.833	0.641	0.933
			Sub-total		47	43	0.915	0.801	0.966
		Sub-total			191	173	0.906	0.856	0.940
		(Pseudo-)replicate 4	Dorsal-flat	0	23	20	0.870	0.679	0.955
			Sub-total		23	20	0.870	0.679	0.955
			Dorsal-oblique	2	24	20	0.833	0.641	0.933
				3	29	27	0.931	0.780	0.981
				4	24	23	0.958	0.798	0.993
				6	27	26	0.963	0.817	0.993
				7	24	23	0.958	0.798	0.993
				8	22	21	0.955	0.782	0.992
			Sub-total		150	140	0.933	0.882	0.963
			Front/back-oblique	1	34	29	0.853	0.699	0.936
				5	24	23	0.958	0.798	0.993
			Sub-total		58	52	0.897	0.792	0.952

		Sub-total			231	212	0.918	0.875	0.947
		(Pseudo-)replicate 5	Dorsal-flat	0	14	12	0.857	0.601	0.960
			Sub-total		14	12	0.857	0.601	0.960
			Dorsal-oblique	2	28	26	0.929	0.774	0.980
				3	21	21	1.000	0.845	1.000
				4	22	20	0.909	0.722	0.975
				6	27	26	0.963	0.817	0.993
				7	12	10	0.833	0.552	0.953
				8	22	20	0.909	0.722	0.975
			Sub-total		132	123	0.932	0.875	0.964
			Front/back-oblique	1	26	24	0.923	0.759	0.979
				5	24	24	1.000	0.862	1.000
			Sub-total		50	48	0.960	0.865	0.989
		Sub-total			196	183	0.934	0.890	0.961
		(Pseudo-)replicate 6	Dorsal-flat	0	24	21	0.875	0.690	0.957
			Sub-total		24	21	0.875	0.690	0.957
			Dorsal-oblique	2	20	19	0.950	0.764	0.991
				3	18	18	1.000	0.824	1.000
				4	18	18	1.000	0.824	1.000
				6	19	18	0.947	0.754	0.991
				7	26	25	0.962	0.811	0.993
				8	14	13	0.929	0.685	0.987
			Sub-total		115	111	0.965	0.914	0.986
			Front/back-oblique	1	25	22	0.880	0.700	0.958
				5	17	13	0.765	0.527	0.904
			Sub-total		42	35	0.833	0.694	0.917
		Sub-total			181	167	0.923	0.874	0.953
		(Pseudo-)replicate 7	Dorsal-flat	0	23	21	0.913	0.732	0.976
			Sub-total		23	21	0.913	0.732	0.976
			Dorsal-oblique	2	19	18	0.947	0.754	0.991
				3	27	25	0.926	0.766	0.979
				4	19	15	0.789	0.567	0.915
				6	28	25	0.893	0.728	0.963
				7	26	24	0.923	0.759	0.979
				8	14	12	0.857	0.601	0.960
			Sub-total		133	119	0.895	0.831	0.936
			Front/back-oblique	1	23	19	0.826	0.629	0.930
				5	20	18	0.900	0.699	0.972

			Sub-total		43	37	0.860	0.727	0.934
		Sub-total			199	177	0.889	0.838	0.926
		(Pseudo-)replicate 8	Dorsal-flat	0	26	25	0.962	0.811	0.993
			Sub-total		26	25	0.962	0.811	0.993
			Dorsal-oblique	2	21	18	0.857	0.654	0.950
				3	23	22	0.957	0.790	0.992
				4	31	29	0.935	0.793	0.982
				6	19	19	1.000	0.832	1.000
				7	25	24	0.960	0.805	0.993
				8	18	18	1.000	0.824	1.000
			Sub-total		137	130	0.949	0.898	0.975
			Front/back-oblique	1	16	15	0.938	0.717	0.989
				5	28	25	0.893	0.728	0.963
			Sub-total		44	40	0.909	0.788	0.964
		Sub-total			207	195	0.942	0.901	0.967
		(Pseudo-)replicate 9	Dorsal-flat	0	23	21	0.913	0.732	0.976
			Sub-total		23	21	0.913	0.732	0.976
			Dorsal-oblique	2	11	9	0.818	0.523	0.949
				3	23	22	0.957	0.790	0.992
				4	32	30	0.938	0.799	0.983
				6	18	18	1.000	0.824	1.000
				7	18	18	1.000	0.824	1.000
				8	27	24	0.889	0.719	0.961
			Sub-total		129	121	0.938	0.882	0.968
			Front/back-oblique	1	19	17	0.895	0.686	0.971
				5	25	21	0.840	0.653	0.936
			Sub-total		44	38	0.864	0.733	0.936
		Sub-total			196	180	0.918	0.872	0.949
		(Pseudo-)replicate 10	Dorsal-flat	0	21	20	0.952	0.773	0.992
			Sub-total		21	20	0.952	0.773	0.992
			Dorsal-oblique	2	23	21	0.913	0.732	0.976
				3	20	19	0.950	0.764	0.991
				4	25	21	0.840	0.653	0.936
				6	15	13	0.867	0.621	0.963
				7	22	18	0.818	0.615	0.927
				8	22	19	0.864	0.667	0.953
			Sub-total		127	111	0.874	0.805	0.921
			Front/back-oblique	1	17	15	0.882	0.657	0.967

				5	19	17	0.895	0.686	0.971
			Sub-total		36	32	0.889	0.747	0.956
		Sub-total			184	163	0.886	0.832	0.924
	Sub-total				1986	1821	0.917	0.904	0.928
	<i>Rhodnius</i>	(Pseudo-)replicate 1	Dorsal-flat	0	36	35	0.972	0.858	0.995
			Sub-total		36	35	0.972	0.858	0.995
			Dorsal-oblique	2	46	46	1.000	0.923	1.000
				3	38	37	0.974	0.865	0.995
				4	46	45	0.978	0.887	0.996
				6	33	33	1.000	0.896	1.000
				7	51	50	0.980	0.897	0.997
				8	45	45	1.000	0.921	1.000
			Sub-total		259	256	0.988	0.967	0.996
			Front/back-oblique	1	47	46	0.979	0.889	0.996
				5	44	44	1.000	0.920	1.000
			Sub-total		91	90	0.989	0.940	0.998
		Sub-total			386	381	0.987	0.970	0.994
		(Pseudo-)replicate 2	Dorsal-flat	0	46	44	0.957	0.855	0.988
			Sub-total		46	44	0.957	0.855	0.988
			Dorsal-oblique	2	43	43	1.000	0.918	1.000
				3	32	32	1.000	0.893	1.000
				4	42	42	1.000	0.916	1.000
				6	44	44	1.000	0.920	1.000
				7	52	52	1.000	0.931	1.000
				8	40	40	1.000	0.912	1.000
			Sub-total		253	253	1.000	0.985	1.000
			Front/back-oblique	1	47	46	0.979	0.889	0.996
				5	34	34	1.000	0.898	1.000
			Sub-total		81	80	0.988	0.933	0.998
		Sub-total			380	377	0.992	0.977	0.997
		(Pseudo-)replicate 3	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	38	38	1.000	0.908	1.000
				3	50	50	1.000	0.929	1.000
				4	44	44	1.000	0.920	1.000
				6	52	52	1.000	0.931	1.000
				7	50	50	1.000	0.929	1.000
				8	43	43	1.000	0.918	1.000

			Sub-total		277	277	1.000	0.986	1.000
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	52	50	0.962	0.870	0.989
			Sub-total		96	94	0.979	0.927	0.994
		Sub-total			414	412	0.995	0.983	0.999
		(Pseudo-)replicate 4	Dorsal-flat	0	43	43	1.000	0.918	1.000
			Sub-total		43	43	1.000	0.918	1.000
			Dorsal-oblique	2	37	37	1.000	0.906	1.000
				3	51	50	0.980	0.897	0.997
				4	35	35	1.000	0.901	1.000
				6	50	49	0.980	0.895	0.996
				7	32	32	1.000	0.893	1.000
				8	42	42	1.000	0.916	1.000
			Sub-total		247	245	0.992	0.971	0.998
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	39	38	0.974	0.868	0.995
			Sub-total		83	82	0.988	0.935	0.998
		Sub-total			373	370	0.992	0.977	0.997
		(Pseudo-)replicate 5	Dorsal-flat	0	50	49	0.980	0.895	0.996
			Sub-total		50	49	0.980	0.895	0.996
			Dorsal-oblique	2	54	53	0.981	0.902	0.997
				3	40	40	1.000	0.912	1.000
				4	50	50	1.000	0.929	1.000
				6	51	50	0.980	0.897	0.997
				7	49	49	1.000	0.927	1.000
				8	38	38	1.000	0.908	1.000
			Sub-total		282	280	0.993	0.975	0.998
			Front/back-oblique	1	35	35	1.000	0.901	1.000
				5	45	45	1.000	0.921	1.000
			Sub-total		80	80	1.000	0.954	1.000
		Sub-total			412	409	0.993	0.979	0.998
		(Pseudo-)replicate 6	Dorsal-flat	0	41	40	0.976	0.874	0.996
			Sub-total		41	40	0.976	0.874	0.996
			Dorsal-oblique	2	43	42	0.977	0.879	0.996
				3	39	39	1.000	0.910	1.000
				4	35	34	0.971	0.855	0.995
				6	40	39	0.975	0.871	0.996
				7	39	39	1.000	0.910	1.000

			8	41	40	0.976	0.874	0.996	
		Sub-total		237	233	0.983	0.957	0.993	
		Front/back-oblique	1	50	49	0.980	0.895	0.996	
			5	47	46	0.979	0.889	0.996	
		Sub-total		97	95	0.979	0.928	0.994	
		Sub-total		375	368	0.981	0.962	0.991	
		(Pseudo-)replicate 7	Dorsal-flat	0	46	45	0.978	0.887	0.996
			Sub-total		46	45	0.978	0.887	0.996
			Dorsal-oblique	2	47	46	0.979	0.889	0.996
				3	45	45	1.000	0.921	1.000
				4	45	44	0.978	0.884	0.996
				6	51	49	0.961	0.868	0.989
				7	34	33	0.971	0.851	0.995
				8	46	46	1.000	0.923	1.000
			Sub-total		268	263	0.981	0.957	0.992
			Front/back-oblique	1	41	41	1.000	0.914	1.000
				5	49	49	1.000	0.927	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total		404	398	0.985	0.968	0.993	
		(Pseudo-)replicate 8	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	52	52	1.000	0.931	1.000
				3	43	43	1.000	0.918	1.000
				4	40	40	1.000	0.912	1.000
				6	37	37	1.000	0.906	1.000
				7	46	46	1.000	0.923	1.000
				8	51	51	1.000	0.930	1.000
			Sub-total		269	269	1.000	0.986	1.000
			Front/back-oblique	1	42	42	1.000	0.916	1.000
				5	48	48	1.000	0.926	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total		400	400	1.000	0.990	1.000	
		(Pseudo-)replicate 9	Dorsal-flat	0	44	42	0.955	0.849	0.987
			Sub-total		44	42	0.955	0.849	0.987
			Dorsal-oblique	2	41	41	1.000	0.914	1.000
				3	56	55	0.982	0.906	0.997
				4	63	60	0.952	0.869	0.984
				6	38	38	1.000	0.908	1.000

				7	40	40	1.000	0.912	1.000
				8	44	44	1.000	0.920	1.000
			Sub-total		282	278	0.986	0.964	0.994
			Front/back-oblique	1	46	45	0.978	0.887	0.996
				5	44	43	0.977	0.882	0.996
			Sub-total		90	88	0.978	0.923	0.994
		Sub-total			416	408	0.981	0.963	0.990
		(Pseudo-)replicate 10	Dorsal-flat	0	45	43	0.956	0.852	0.988
			Sub-total		45	43	0.956	0.852	0.988
			Dorsal-oblique	2	44	44	1.000	0.920	1.000
				3	37	37	1.000	0.906	1.000
				4	40	40	1.000	0.912	1.000
				6	45	44	0.978	0.884	0.996
				7	44	43	0.977	0.882	0.996
				8	42	42	1.000	0.916	1.000
			Sub-total		252	250	0.992	0.972	0.998
			Front/back-oblique	1	52	51	0.981	0.899	0.997
				5	52	51	0.981	0.899	0.997
			Sub-total		104	102	0.981	0.933	0.995
		Sub-total			401	395	0.985	0.968	0.993
	Sub-total				3961	3918	0.989	0.985	0.992
Total HB					13,140	12,825	0.976	0.973	0.979
Linear discriminant model (LD)	<i>Triatoma</i>	(Pseudo-)replicate 1	Dorsal-flat	0	84	82	0.976	0.917	0.993
			Sub-total		84	82	0.976	0.917	0.993
			Dorsal-oblique	2	81	78	0.963	0.897	0.987
				3	88	85	0.966	0.905	0.988
				4	92	89	0.967	0.908	0.989
				6	80	77	0.963	0.895	0.987
				7	83	80	0.964	0.899	0.988
				8	78	76	0.974	0.911	0.993
			Sub-total		502	485	0.966	0.946	0.979
			Front/back-oblique	1	68	68	1.000	0.947	1.000
				5	72	70	0.972	0.904	0.992
			Sub-total		140	138	0.986	0.949	0.996
		Sub-total			726	705	0.971	0.956	0.981
		(Pseudo-)replicate 2	Dorsal-flat	0	72	69	0.958	0.885	0.986
			Sub-total		72	69	0.958	0.885	0.986
			Dorsal-oblique	2	89	86	0.966	0.906	0.988

				3	73	72	0.986	0.926	0.998
				4	92	88	0.957	0.893	0.983
				6	77	76	0.987	0.930	0.998
				7	78	75	0.962	0.893	0.987
				8	88	86	0.977	0.921	0.994
			Sub-total		497	483	0.972	0.953	0.983
			Front/back-oblique	1	83	80	0.964	0.899	0.988
				5	83	80	0.964	0.899	0.988
			Sub-total		166	160	0.964	0.923	0.983
		Sub-total			735	712	0.969	0.953	0.979
		(Pseudo-)replicate 3	Dorsal-flat	0	72	69	0.958	0.885	0.986
			Sub-total		72	69	0.958	0.885	0.986
			Dorsal-oblique	2	75	70	0.933	0.853	0.971
				3	77	73	0.948	0.874	0.980
				4	90	88	0.978	0.923	0.994
				6	83	81	0.976	0.916	0.993
				7	71	66	0.930	0.846	0.970
				8	65	62	0.954	0.873	0.984
			Sub-total		461	440	0.954	0.931	0.970
			Front/back-oblique	1	83	82	0.988	0.935	0.998
				5	93	90	0.968	0.909	0.989
			Sub-total		176	172	0.977	0.943	0.991
		Sub-total			709	681	0.961	0.944	0.973
		(Pseudo-)replicate 4	Dorsal-flat	0	78	75	0.962	0.893	0.987
			Sub-total		78	75	0.962	0.893	0.987
			Dorsal-oblique	2	78	73	0.936	0.859	0.972
				3	87	85	0.977	0.920	0.994
				4	75	72	0.960	0.889	0.986
				6	82	81	0.988	0.934	0.998
				7	77	77	1.000	0.952	1.000
				8	84	81	0.964	0.900	0.988
			Sub-total		483	469	0.971	0.952	0.983
			Front/back-oblique	1	75	74	0.987	0.928	0.998
				5	74	73	0.986	0.927	0.998
			Sub-total		149	147	0.987	0.952	0.996
		Sub-total			710	691	0.973	0.959	0.983
		(Pseudo-)replicate 5	Dorsal-flat	0	73	71	0.973	0.905	0.992
			Sub-total		73	71	0.973	0.905	0.992

			Dorsal-oblique	2	71	67	0.944	0.864	0.978
				3	82	77	0.939	0.865	0.974
				4	88	83	0.943	0.874	0.975
				6	78	76	0.974	0.911	0.993
				7	100	95	0.950	0.888	0.978
				8	76	73	0.961	0.890	0.986
			Sub-total		495	471	0.952	0.929	0.967
			Front/back-oblique	1	72	72	1.000	0.949	1.000
				5	66	64	0.970	0.896	0.992
			Sub-total		138	136	0.986	0.949	0.996
		Sub-total			706	678	0.960	0.943	0.972
		(Pseudo-)replicate 6	Dorsal-flat	0	80	79	0.988	0.933	0.998
			Sub-total		80	79	0.988	0.933	0.998
			Dorsal-oblique	2	76	74	0.974	0.909	0.993
				3	91	90	0.989	0.940	0.998
				4	80	77	0.963	0.895	0.987
				6	80	79	0.988	0.933	0.998
				7	85	82	0.965	0.901	0.988
				8	81	78	0.963	0.897	0.987
			Sub-total		493	480	0.974	0.955	0.985
			Front/back-oblique	1	96	94	0.979	0.927	0.994
				5	89	88	0.989	0.939	0.998
			Sub-total		185	182	0.984	0.953	0.994
		Sub-total			758	741	0.978	0.964	0.986
		(Pseudo-)replicate 7	Dorsal-flat	0	87	83	0.954	0.888	0.982
			Sub-total		87	83	0.954	0.888	0.982
			Dorsal-oblique	2	74	72	0.973	0.907	0.993
				3	87	86	0.989	0.938	0.998
				4	76	75	0.987	0.929	0.998
				6	73	72	0.986	0.926	0.998
				7	77	74	0.961	0.892	0.987
				8	86	82	0.953	0.886	0.982
			Sub-total		473	461	0.975	0.956	0.985
			Front/back-oblique	1	82	81	0.988	0.934	0.998
				5	69	69	1.000	0.947	1.000
			Sub-total		151	150	0.993	0.963	0.999
		Sub-total			711	694	0.976	0.962	0.985
		(Pseudo-)replicate 8	Dorsal-flat	0	91	88	0.967	0.908	0.989

			Sub-total		91	88	0.967	0.908	0.989
			Dorsal-oblique	2	71	69	0.972	0.903	0.992
				3	84	83	0.988	0.936	0.998
				4	66	62	0.939	0.854	0.976
				6	78	76	0.974	0.911	0.993
				7	80	74	0.925	0.846	0.965
				8	62	57	0.919	0.825	0.965
			Sub-total		441	421	0.955	0.931	0.970
			Front/back-oblique	1	77	75	0.974	0.910	0.993
				5	98	98	1.000	0.962	1.000
			Sub-total		175	173	0.989	0.959	0.997
		Sub-total			707	682	0.965	0.948	0.976
		(Pseudo-)replicate 9	Dorsal-flat	0	79	78	0.987	0.932	0.998
			Sub-total		79	78	0.987	0.932	0.998
			Dorsal-oblique	2	90	89	0.989	0.940	0.998
				3	85	83	0.976	0.918	0.994
				4	96	91	0.948	0.884	0.978
				6	84	81	0.964	0.900	0.988
				7	63	62	0.984	0.915	0.997
				8	69	67	0.971	0.900	0.992
			Sub-total		487	473	0.971	0.952	0.983
			Front/back-oblique	1	64	61	0.953	0.871	0.984
				5	72	70	0.972	0.904	0.992
			Sub-total		136	131	0.963	0.917	0.984
		Sub-total			702	682	0.972	0.956	0.981
		(Pseudo-)replicate 10	Dorsal-flat	0	71	71	1.000	0.949	1.000
			Sub-total		71	71	1.000	0.949	1.000
			Dorsal-oblique	2	73	72	0.986	0.926	0.998
				3	81	78	0.963	0.897	0.987
				4	83	83	1.000	0.956	1.000
				6	103	101	0.981	0.932	0.995
				7	63	60	0.952	0.869	0.984
				8	84	82	0.976	0.917	0.993
			Sub-total		487	476	0.977	0.960	0.987
			Front/back-oblique	1	76	76	1.000	0.952	1.000
				5	95	93	0.979	0.926	0.994
			Sub-total		171	169	0.988	0.958	0.997
		Sub-total			729	716	0.982	0.970	0.990

	Sub-total				7193	6982	0.971	0.967	0.974
	<i>Panstrongylus</i>	(Pseudo-)replicate 1	Dorsal-flat	0	20	15	0.750	0.531	0.888
			Sub-total		20	15	0.750	0.531	0.888
			Dorsal-oblique	2	22	21	0.955	0.782	0.992
				3	25	23	0.920	0.750	0.978
				4	20	16	0.800	0.584	0.919
				6	23	20	0.870	0.679	0.955
				7	25	22	0.880	0.700	0.958
				8	29	26	0.897	0.736	0.964
			Sub-total		144	128	0.889	0.827	0.930
			Front/back-oblique	1	20	18	0.900	0.699	0.972
				5	18	13	0.722	0.491	0.875
			Sub-total		38	31	0.816	0.666	0.908
		Sub-total			202	174	0.861	0.807	0.902
		(Pseudo-)replicate 2	Dorsal-flat	0	22	17	0.773	0.566	0.899
			Sub-total		22	17	0.773	0.566	0.899
			Dorsal-oblique	2	20	17	0.850	0.640	0.948
				3	17	17	1.000	0.816	1.000
				4	27	22	0.815	0.633	0.918
				6	23	20	0.870	0.679	0.955
				7	22	19	0.864	0.667	0.953
				8	23	19	0.826	0.629	0.930
			Sub-total		132	114	0.864	0.795	0.912
			Front/back-oblique	1	20	19	0.950	0.764	0.991
				5	25	23	0.920	0.750	0.978
			Sub-total		45	42	0.933	0.821	0.977
		Sub-total			199	173	0.869	0.815	0.909
		(Pseudo-)replicate 3	Dorsal-flat	0	22	20	0.909	0.722	0.975
			Sub-total		22	20	0.909	0.722	0.975
			Dorsal-oblique	2	17	15	0.882	0.657	0.967
				3	17	16	0.941	0.730	0.990
				4	17	14	0.824	0.590	0.938
				6	20	15	0.750	0.531	0.888
				7	24	20	0.833	0.641	0.933
				8	27	23	0.852	0.675	0.941
			Sub-total		122	103	0.844	0.770	0.898
			Front/back-oblique	1	23	22	0.957	0.790	0.992
				5	24	22	0.917	0.742	0.977

			Sub-total		47	44	0.936	0.828	0.978
		Sub-total			191	167	0.874	0.820	0.914
		(Pseudo-)replicate 4	Dorsal-flat	0	23	20	0.870	0.679	0.955
			Sub-total		23	20	0.870	0.679	0.955
			Dorsal-oblique	2	24	19	0.792	0.595	0.908
				3	29	25	0.862	0.694	0.945
				4	24	19	0.792	0.595	0.908
				6	27	25	0.926	0.766	0.979
				7	24	22	0.917	0.742	0.977
				8	22	21	0.955	0.782	0.992
			Sub-total		150	131	0.873	0.811	0.917
			Front/back-oblique	1	34	29	0.853	0.699	0.936
				5	24	21	0.875	0.690	0.957
			Sub-total		58	50	0.862	0.751	0.928
		Sub-total			231	201	0.870	0.821	0.907
		(Pseudo-)replicate 5	Dorsal-flat	0	14	13	0.929	0.685	0.987
			Sub-total		14	13	0.929	0.685	0.987
			Dorsal-oblique	2	28	25	0.893	0.728	0.963
				3	21	21	1.000	0.845	1.000
				4	22	20	0.909	0.722	0.975
				6	27	23	0.852	0.675	0.941
				7	12	9	0.750	0.468	0.911
				8	22	19	0.864	0.667	0.953
			Sub-total		132	117	0.886	0.821	0.930
			Front/back-oblique	1	26	24	0.923	0.759	0.979
				5	24	22	0.917	0.742	0.977
			Sub-total		50	46	0.920	0.812	0.968
		Sub-total			196	176	0.898	0.848	0.933
		(Pseudo-)replicate 6	Dorsal-flat	0	24	20	0.833	0.641	0.933
			Sub-total		24	20	0.833	0.641	0.933
			Dorsal-oblique	2	20	18	0.900	0.699	0.972
				3	18	18	1.000	0.824	1.000
				4	18	17	0.944	0.742	0.990
				6	19	17	0.895	0.686	0.971
				7	26	21	0.808	0.621	0.915
				8	14	11	0.786	0.524	0.924
			Sub-total		115	102	0.887	0.816	0.933
			Front/back-oblique	1	25	20	0.800	0.609	0.911

			5	17	12	0.706	0.469	0.867
		Sub-total		42	32	0.762	0.615	0.865
		Sub-total		181	154	0.851	0.792	0.895
	(Pseudo-)replicate 7	Dorsal-flat	0	23	21	0.913	0.732	0.976
		Sub-total		23	21	0.913	0.732	0.976
		Dorsal-oblique	2	19	19	1.000	0.832	1.000
			3	27	25	0.926	0.766	0.979
			4	19	16	0.842	0.624	0.945
			6	28	24	0.857	0.685	0.943
			7	26	20	0.769	0.579	0.890
			8	14	8	0.571	0.326	0.786
		Sub-total		133	112	0.842	0.771	0.894
		Front/back-oblique	1	23	21	0.913	0.732	0.976
			5	20	16	0.800	0.584	0.919
		Sub-total		43	37	0.860	0.727	0.934
		Sub-total		199	170	0.854	0.799	0.897
	(Pseudo-)replicate 8	Dorsal-flat	0	26	23	0.885	0.710	0.960
		Sub-total		26	23	0.885	0.710	0.960
		Dorsal-oblique	2	21	21	1.000	0.845	1.000
			3	23	20	0.870	0.679	0.955
			4	31	26	0.839	0.674	0.929
			6	19	18	0.947	0.754	0.991
			7	25	23	0.920	0.750	0.978
			8	18	17	0.944	0.742	0.990
		Sub-total		137	125	0.912	0.853	0.949
		Front/back-oblique	1	16	14	0.875	0.640	0.965
			5	28	24	0.857	0.685	0.943
		Sub-total		44	38	0.864	0.733	0.936
		Sub-total		207	186	0.899	0.850	0.933
	(Pseudo-)replicate 9	Dorsal-flat	0	23	21	0.913	0.732	0.976
		Sub-total		23	21	0.913	0.732	0.976
		Dorsal-oblique	2	11	8	0.727	0.434	0.903
			3	23	20	0.870	0.679	0.955
			4	32	31	0.969	0.843	0.994
			6	18	16	0.889	0.672	0.969
			7	18	16	0.889	0.672	0.969
			8	27	24	0.889	0.719	0.961
		Sub-total		129	115	0.891	0.826	0.934

			Front/back-oblique	1	19	15	0.789	0.567	0.915
				5	25	20	0.800	0.609	0.911
			Sub-total		44	35	0.795	0.655	0.888
		Sub-total			196	171	0.872	0.818	0.912
		(Pseudo-)replicate 10	Dorsal-flat	0	21	18	0.857	0.654	0.950
			Sub-total		21	18	0.857	0.654	0.950
			Dorsal-oblique	2	23	21	0.913	0.732	0.976
				3	20	20	1.000	0.839	1.000
				4	25	20	0.800	0.609	0.911
				6	15	14	0.933	0.702	0.988
				7	22	17	0.773	0.566	0.899
				8	22	16	0.727	0.518	0.868
			Sub-total		127	108	0.850	0.778	0.902
			Front/back-oblique	1	17	15	0.882	0.657	0.967
				5	19	17	0.895	0.686	0.971
			Sub-total		36	32	0.889	0.747	0.956
		Sub-total			184	158	0.859	0.801	0.902
	Sub-total				1986	1730	0.871	0.856	0.885
	<i>Rhodnius</i>	(Pseudo-)replicate 1	Dorsal-flat	0	36	34	0.944	0.819	0.985
			Sub-total		36	34	0.944	0.819	0.985
			Dorsal-oblique	2	46	45	0.978	0.887	0.996
				3	38	37	0.974	0.865	0.995
				4	46	45	0.978	0.887	0.996
				6	33	31	0.939	0.804	0.983
				7	51	51	1.000	0.930	1.000
				8	45	44	0.978	0.884	0.996
			Sub-total		259	253	0.977	0.950	0.989
			Front/back-oblique	1	47	46	0.979	0.889	0.996
				5	44	42	0.955	0.849	0.987
			Sub-total		91	88	0.967	0.908	0.989
		Sub-total			386	375	0.972	0.950	0.984
		(Pseudo-)replicate 2	Dorsal-flat	0	46	45	0.978	0.887	0.996
			Sub-total		46	45	0.978	0.887	0.996
			Dorsal-oblique	2	43	43	1.000	0.918	1.000
				3	32	30	0.938	0.799	0.983
				4	42	42	1.000	0.916	1.000
				6	44	44	1.000	0.920	1.000
				7	52	50	0.962	0.870	0.989

			8	40	40	1.000	0.912	1.000	
		Sub-total		253	249	0.984	0.960	0.994	
		Front/back-oblique	1	47	46	0.979	0.889	0.996	
			5	34	32	0.941	0.809	0.984	
		Sub-total		81	78	0.963	0.897	0.987	
		Sub-total		380	372	0.979	0.959	0.989	
		(Pseudo-)replicate 3	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	38	37	0.974	0.865	0.995
				3	50	50	1.000	0.929	1.000
				4	44	44	1.000	0.920	1.000
				6	52	52	1.000	0.931	1.000
				7	50	49	0.980	0.895	0.996
				8	43	42	0.977	0.879	0.996
			Sub-total		277	274	0.989	0.969	0.996
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	52	52	1.000	0.931	1.000
			Sub-total		96	96	1.000	0.962	1.000
		Sub-total		414	411	0.993	0.979	0.998	
		(Pseudo-)replicate 4	Dorsal-flat	0	43	41	0.953	0.845	0.987
			Sub-total		43	41	0.953	0.845	0.987
			Dorsal-oblique	2	37	36	0.973	0.862	0.995
				3	51	50	0.980	0.897	0.997
				4	35	35	1.000	0.901	1.000
				6	50	47	0.940	0.838	0.979
				7	32	32	1.000	0.893	1.000
				8	42	42	1.000	0.916	1.000
			Sub-total		247	242	0.980	0.953	0.991
			Front/back-oblique	1	44	44	1.000	0.920	1.000
				5	39	39	1.000	0.910	1.000
			Sub-total		83	83	1.000	0.956	1.000
		Sub-total		373	366	0.981	0.962	0.991	
		(Pseudo-)replicate 5	Dorsal-flat	0	50	50	1.000	0.929	1.000
			Sub-total		50	50	1.000	0.929	1.000
			Dorsal-oblique	2	54	52	0.963	0.875	0.990
				3	40	40	1.000	0.912	1.000
				4	50	50	1.000	0.929	1.000
				6	51	48	0.941	0.841	0.980

				7	49	49	1.000	0.927	1.000
				8	38	38	1.000	0.908	1.000
			Sub-total		282	277	0.982	0.959	0.992
			Front/back-oblique	1	35	34	0.971	0.855	0.995
				5	45	45	1.000	0.921	1.000
			Sub-total		80	79	0.988	0.933	0.998
		Sub-total			412	406	0.985	0.969	0.993
		(Pseudo-)replicate 6	Dorsal-flat	0	41	41	1.000	0.914	1.000
			Sub-total		41	41	1.000	0.914	1.000
			Dorsal-oblique	2	43	41	0.953	0.845	0.987
				3	39	39	1.000	0.910	1.000
				4	35	34	0.971	0.855	0.995
				6	40	39	0.975	0.871	0.996
				7	39	38	0.974	0.868	0.995
				8	41	39	0.951	0.839	0.987
			Sub-total		237	230	0.970	0.940	0.986
			Front/back-oblique	1	50	49	0.980	0.895	0.996
				5	47	47	1.000	0.924	1.000
			Sub-total		97	96	0.990	0.944	0.998
		Sub-total			375	367	0.979	0.958	0.989
		(Pseudo-)replicate 7	Dorsal-flat	0	46	45	0.978	0.887	0.996
			Sub-total		46	45	0.978	0.887	0.996
			Dorsal-oblique	2	47	46	0.979	0.889	0.996
				3	45	44	0.978	0.884	0.996
				4	45	45	1.000	0.921	1.000
				6	51	49	0.961	0.868	0.989
				7	34	34	1.000	0.898	1.000
				8	46	46	1.000	0.923	1.000
			Sub-total		268	264	0.985	0.962	0.994
			Front/back-oblique	1	41	41	1.000	0.914	1.000
				5	49	49	1.000	0.927	1.000
			Sub-total		90	90	1.000	0.959	1.000
		Sub-total			404	399	0.988	0.971	0.995
		(Pseudo-)replicate 8	Dorsal-flat	0	41	40	0.976	0.874	0.996
			Sub-total		41	40	0.976	0.874	0.996
			Dorsal-oblique	2	52	51	0.981	0.899	0.997
				3	43	42	0.977	0.879	0.996
				4	40	39	0.975	0.871	0.996

				6	37	36	0.973	0.862	0.995
				7	46	45	0.978	0.887	0.996
				8	51	51	1.000	0.930	1.000
			Sub-total		269	264	0.981	0.957	0.992
			Front/back-oblique	1	42	41	0.976	0.877	0.996
				5	48	48	1.000	0.926	1.000
			Sub-total		90	89	0.989	0.940	0.998
		Sub-total			400	393	0.983	0.964	0.991
		(Pseudo-)replicate 9	Dorsal-flat	0	44	42	0.955	0.849	0.987
			Sub-total		44	42	0.955	0.849	0.987
			Dorsal-oblique	2	41	41	1.000	0.914	1.000
				3	56	56	1.000	0.936	1.000
				4	63	62	0.984	0.915	0.997
				6	38	37	0.974	0.865	0.995
				7	40	38	0.950	0.835	0.986
				8	44	44	1.000	0.920	1.000
			Sub-total		282	278	0.986	0.964	0.994
			Front/back-oblique	1	46	46	1.000	0.923	1.000
				5	44	43	0.977	0.882	0.996
			Sub-total		90	89	0.989	0.940	0.998
		Sub-total			416	409	0.983	0.966	0.992
		(Pseudo-)replicate 10	Dorsal-flat	0	45	43	0.956	0.852	0.988
			Sub-total		45	43	0.956	0.852	0.988
			Dorsal-oblique	2	44	44	1.000	0.920	1.000
				3	37	36	0.973	0.862	0.995
				4	40	38	0.950	0.835	0.986
				6	45	44	0.978	0.884	0.996
				7	44	43	0.977	0.882	0.996
				8	42	41	0.976	0.877	0.996
			Sub-total		252	246	0.976	0.949	0.989
			Front/back-oblique	1	52	50	0.962	0.870	0.989
				5	52	51	0.981	0.899	0.997
			Sub-total		104	101	0.971	0.919	0.990
		Sub-total			401	390	0.973	0.952	0.985
	Sub-total				3961	3888	0.982	0.977	0.985
Total LD					13,140	12,600	0.959	0.955	0.962
Grand total					65,700	60,329	0.918	0.916	0.920

Angle, each of 9 angles at which bugs were pictured (see Fig 1 in the main text); (Pseudo-)replicate: each of the 10 (pseudo-)replicate training-testing runs each algorithm completed; all 5 algorithms used the same training and testing datasets in each (pseudo-)replicate run; “N”, number of pictures used in (pseudo-)replicate training-testing runs of the 5 algorithms. The Grand total (65,700) is the total number of observations (rows in the full database) used for performance analyses; it corresponds to 10 random subsets of 1314 pictures (20% of the 9×730 specimens = 6570 unique pictures available) used in each of the 10 (pseudo-)replicate testing runs completed by 5 algorithms ($1314 \times 10 \times 5 = 65,700$ observations). The Total value for each algorithm corresponds to 1314 pictures \times 10 (pseudo-)replicates = 13,140 observations; Correct: the number of pictures that were correctly identified at the genus level; CI lower and CI upper: lower and upper limits of the score 95% confidence interval for each proportion

Table S7. Top-ranking accuracy model ('A×G+P'): estimates, standard errors (SE), and 95% confidence interval limits.

Term	Estimate	SE	CI lower	CI upper
Fixed effects*				
Intercept	12.048	0.577	10.916	13.180
Algorithm AB (AdaBoost)	-7.326	0.544	-8.393	-6.259
Algorithm GB (Gradient Boosting)	-5.203	0.546	-6.274	-3.133
Algorithm HB (Histogram Boosting)	-4.432	0.549	-5.508	-3.357
Algorithm LD (Linear Discriminant)	-5.398	0.546	-6.468	-4.329
Genus <i>Panstrongylus</i>	-1.033	0.928	-2.852	0.786
Genus <i>Rhodnius</i>	-0.317	0.957	-2.193	1.559
AB × <i>Panstrongylus</i>	-2.701	0.908	-4.481	-0.920
GB × <i>Panstrongylus</i>	-1.694	0.910	-3.477	0.090
HB × <i>Panstrongylus</i>	-1.967	0.912	-3.755	-0.197
LD × <i>Panstrongylus</i>	-1.829	0.909	-3.611	-0.046
AB × <i>Rhodnius</i>	-5.339	0.949	-7.199	-3.479
GB × <i>Rhodnius</i>	0.003	0.952	-1.862	1.868
HB × <i>Rhodnius</i>	-0.284	0.956	-2.157	1.591
LD × <i>Rhodnius</i>	-0.057	0.951	-1.921	1.806
Position B (dorsal-oblique)	-0.306	0.120	-0.541	-0.072
Position C (front/back-oblique)	-0.182	0.136	-0.449	0.084
Random effects (variance)				
Picture (nested within bug)	2.546	-	-	-
Bug	3.683	-	-	-
Pseudo-replicate	0.065	-	-	-

*Reference levels are: AN (AlexNet) for algorithm, *Triatoma* for genus, and A (dorsal-flat) for position

Table S8. Data underlying Fig. 4: predicted accuracy, top-ranking model ('A×G+P'), with 95% confidence interval limits.

Algorithm	Genus	Position	Accuracy	CI lower	CI upper
AlexNet	<i>Triatoma</i>	A (dorsal-flat)	0.999994	0.999982	0.999998
		B (dorsal-oblique)	0.999992	0.999976	0.999997
		C (front/back-oblique)	0.999993	0.999978	0.999998
	<i>Panstrongylus</i>	A	0.999984	0.999902	0.999997
		B	0.999978	0.999869	0.999996
		C	0.999980	0.999884	0.999997
	<i>Rhodnius</i>	A	0.999992	0.999947	0.999999
		B	0.999989	0.999929	0.999998
		C	0.999990	0.999937	0.999999
AdaBoost	<i>Triatoma</i>	A	0.991183	0.987045	0.994008
		B	0.988062	0.983468	0.991391
		C	0.989439	0.985009	0.992569
	<i>Panstrongylus</i>	A	0.728850	0.625200	0.812436
		B	0.664311	0.561946	0.753259
		C	0.691360	0.587799	0.778698
	<i>Rhodnius</i>	A	0.282099	0.211354	0.365547
		B	0.224382	0.172388	0.286627
		C	0.246682	0.187333	0.317487
Gradient Boosting	<i>Triatoma</i>	A	0.998936	0.998359	0.999310
		B	0.998555	0.997887	0.999012
		C	0.998723	0.998091	0.999146
	<i>Panstrongylus</i>	A	0.983980	0.973565	0.990332
		B	0.978364	0.965823	0.986368
		C	0.980838	0.969176	0.988141
	<i>Rhodnius</i>	A	0.998543	0.997549	0.999135
		B	0.998022	0.996811	0.998774
		C	0.998253	0.997135	0.998934
Histogram Boosting	<i>Triatoma</i>	A	0.999507	0.999219	0.999689
		B	0.999331	0.998991	0.999557
		C	0.999409	0.999090	0.999616
	<i>Panstrongylus</i>	A	0.990198	0.983573	0.994168
		B	0.986733	0.978680	0.991770
		C	0.988261	0.980805	0.992842
	<i>Rhodnius</i>	A	0.999102	0.998429	0.999487
		B	0.998781	0.997949	0.999276
		C	0.998923	0.998160	0.999370
Linear Discriminant	<i>Triatoma</i>	A	0.998707	0.998018	0.999157
		B	0.998245	0.997449	0.998793
		C	0.998449	0.997695	0.998957
	<i>Panstrongylus</i>	A	0.977859	0.963834	0.986521
		B	0.970162	0.953398	0.981016
		C	0.973548	0.957907	0.983477
	<i>Rhodnius</i>	A	0.998121	0.996893	0.998864
		B	0.997449	0.995966	0.998388
		C	0.997746	0.996374	0.998599

Table S9. Data underlying Fig. 5: observed specificity by algorithm and genus (with score 95% confidence interval limits).

Genus	Algorithm	“N”	Correct	Proportion	CI lower	CI upper
<i>Triatoma</i>	AlexNet	5947	5947	1.000000	0.999354	1.000000
	AdaBoost	5947	2564	0.431142	0.418604	0.443769
	Gradient Boosting	5947	5670	0.953422	0.947767	0.958491
	Histogram Boosting	5947	5739	0.965024	0.960047	0.969402
	Linear Discriminant	5947	5618	0.944678	0.938576	0.950206
<i>Panstrongylus</i>	AlexNet	11154	11154	1.000000	0.999656	1.000000
	AdaBoost	11154	7906	0.708804	0.700302	0.717162
	Gradient Boosting	11154	10908	0.977945	0.975051	0.980511
	Histogram Boosting	11154	11004	0.986552	0.984241	0.988528
	Linear Discriminant	11154	10870	0.974538	0.971448	0.977302
<i>Rhodnius</i>	AlexNet	9179	9179	1.000000	0.999582	1.000000
	AdaBoost	9179	7702	0.839089	0.831431	0.846464
	Gradient Boosting	9179	8778	0.956313	0.951938	0.960307
	Histogram Boosting	9179	8907	0.970367	0.966697	0.973644
	Linear Discriminant	9179	8712	0.949123	0.944437	0.953434

Table S10. Top-ranking *Triatoma* specificity model (‘T A+G+P’): estimates, standard errors (SE), and 95% confidence intervals.

Term	Estimate	SE	CI lower	CI upper
Fixed effects*				
Intercept	7.707	0.577	6.576	8.838
Algorithm AB (AdaBoost)	-7.935	0.549	-9.012	-6.858
Algorithm GB (Gradient Boosting)	-4.509	0.552	-5.591	-3.427
Algorithm HB (Histogram Boosting)	-4.207	0.553	-5.292	-3.123
Algorithm LD (Linear Discriminant)	-4.692	0.552	-5.773	-3.611
Genus <i>Rhodnius</i>	0.096	0.046	0.006	0.186
Position B (dorsal-oblique)	-0.168	0.072	-0.308	-0.027
Position C (front/back-oblique)	-0.167	0.081	-0.325	-0.009
Random effects (variance)				
Pseudo-replicate	0.272	-	-	-

*Reference levels are: AN (AlexNet) for algorithm, *Triatoma* for genus, and A (dorsal-flat) for position

Table S11. Top-ranking *Panstrongylus* specificity model (‘P A+G+P’): estimates, standard errors (SE), and 95% confidence intervals.

Term	Estimate	SE	CI lower	CI upper
Fixed effects*				
Intercept	8.041	0.573	6.919	9.163
Algorithm AB (AdaBoost)	-5.549	0.544	-6.615	-4.483
Algorithm GB (Gradient Boosting)	-3.706	0.550	-4.784	-2.628
Algorithm HB (Histogram Boosting)	-3.216	0.555	-4.304	-2.128
Algorithm LD (Linear Discriminant)	-4.172	0.547	-5.245	-3.099
Genus <i>Rhodnius</i>	6.666	1.076	4.557	8.775
Position B (dorsal-oblique)	-0.113	0.110	-0.329	0.107
Position C (front/back-oblique)	0.109	0.128	-0.141	0.360
Random effects (variance)				
Pseudo-replicate	0.282	-	-	-

*Reference levels are: AN (AlexNet) for algorithm, *Triatoma* for genus, and A (dorsal-flat) for position

Table S12. Top-ranking *Rhodnius* specificity model ('R A+G+P'): estimates, standard errors (SE), and 95% confidence intervals (CI).

Term	Estimate	SE	CI lower	CI upper
Fixed effects*				
Intercept	8.503	0.655	7.220	9.786
Algorithm AB (AdaBoost)	-2.933	0.622	-4.152	-1.714
Algorithm GB (Gradient Boosting)	-3.496	0.614	-4.670	-2.293
Algorithm HB (Histogram Boosting)	-2.834	0.623	-4.055	-1.613
Algorithm LD (Linear Discriminant)	-2.808	0.624	-4.031	-1.585
Genus <i>Panstrongylus</i>	4.656	1.270	2.168	7.145
Position B (dorsal-oblique)	-0.594	0.277	-1.137	-0.051
Position C (front/back-oblique)	0.005	0.325	-0.632	0.642
Random effects (variance)				
Pseudo-replicate	0.034	-	-	-

*Reference levels are: AN (AlexNet) for algorithm, *Triatoma* for genus, and A (dorsal-flat) for position

Table S13. Data underlying Fig. 6: predicted specificity, top-ranking *Triatoma* model ('T A+G+P'), with 95% confidence interval limits.

Algorithm	Genus	Position	Specificity	CI lower	CI upper
AlexNet	<i>Panstrongylus</i>	A (dorsal-flat)	0.999551	0.998609	0.999855
		B (dorsal-oblique)	0.999468	0.998365	0.999827
		C (front/back-oblique)	0.999469	0.998362	0.999828
	<i>Rhodnius</i>	A	0.999592	0.998737	0.999868
		B	0.999517	0.998516	0.999843
		C	0.999517	0.998513	0.999843
AdaBoost	<i>Panstrongylus</i>	A	0.443206	0.358075	0.531809
		B	0.402283	0.324947	0.484806
		C	0.402400	0.323358	0.486860
	<i>Rhodnius</i>	A	0.466983	0.381307	0.554651
		B	0.425542	0.347262	0.507742
		C	0.425661	0.345611	0.509807
Gradient Boosting	<i>Panstrongylus</i>	A	0.960763	0.944039	0.972634
		B	0.953924	0.935718	0.967155
		C	0.953945	0.935303	0.967403
	<i>Rhodnius</i>	A	0.964222	0.949046	0.974997
		B	0.957960	0.941438	0.969970
		C	0.957980	0.941056	0.970198
Histogram Boosting	<i>Panstrongylus</i>	A	0.970675	0.957743	0.979733
		B	0.965502	0.951340	0.975648
		C	0.965518	0.951026	0.975831
	<i>Rhodnius</i>	A	0.973285	0.961572	0.981497
		B	0.968558	0.955732	0.977754
		C	0.968572	0.955445	0.977922
Linear Discriminant	<i>Panstrongylus</i>	A	0.953239	0.933712	0.967215
		B	0.945164	0.923991	0.960690
		C	0.945189	0.923501	0.960988
	<i>Rhodnius</i>	A	0.957332	0.939586	0.970032
		B	0.949927	0.930679	0.964037
		C	0.949950	0.930228	0.964312

Table S14. Data underlying Fig. 6: predicted specificity, top-ranking *Panstrongylus* model ('P A+G+P'), with 95% confidence interval limits.

Algorithm	Genus	Position	Specificity	CI lower	CI upper	
AlexNet	<i>Triatoma</i>	A (dorsal-flat)	0.9996781	0.9990117	0.9998952	
		B (dorsal-oblique)	0.9996395	0.9989076	0.9998811	
		C (front/back-oblique)	0.9998946	0.9991023	0.9999876	
	<i>Rhodnius</i>	A	0.9999996	0.9999956	1.0000000	
		B	0.9999995	0.9999951	1.0000000	
		C	0.9999996	0.9999961	1.0000000	
	AdaBoost	<i>Triatoma</i>	A	0.9151782	0.8844804	0.9382876
			B	0.9309198	0.9033600	0.9510460
			C	0.9997113	0.9991184	0.9999055
<i>Rhodnius</i>		A	0.9998820	0.9990028	0.9999860	
		B	0.9999055	0.9991985	0.9999889	
		C	0.9235611	0.8911193	0.9469124	
Gradient Boosting		<i>Triatoma</i>	A	0.9870620	0.9803365	0.9915071
			B	0.9855341	0.9789243	0.9900919
			C	0.9883845	0.9827030	0.9922145
	<i>Rhodnius</i>	A	0.9999833	0.9998567	0.9999981	
		B	0.9999813	0.9998408	0.9999978	
		C	0.9999850	0.9998721	0.9999982	
	Histogram Boosting	<i>Triatoma</i>	A	0.9920357	0.9875634	0.9949080
			B	0.9910898	0.9866293	0.9940713
			C	0.9928534	0.9890559	0.9953394
<i>Rhodnius</i>		A	0.9999898	0.9999118	0.9999988	
		B	0.9999886	0.9999020	0.9999987	
		C	0.9999908	0.9999212	0.9999989	
Linear Discriminant		<i>Triatoma</i>	A	0.9795503	0.9695020	0.9863346
			B	0.9771557	0.9673918	0.9840443
			C	0.9816262	0.9731584	0.9874571
	<i>Rhodnius</i>	A	0.9999734	0.9997725	0.9999969	
		B	0.9999702	0.9997473	0.9999965	
		C	0.9999762	0.9997969	0.9999972	

Table S15. Data underlying Fig. 6: predicted specificity, top-ranking *Rhodnius* model ('R A+G+P'), with 95% confidence interval limits.

Algorithm	Genus	Position	Specificity	CI lower	CI upper
AlexNet	<i>Triatoma</i>	A (dorsal-flat)	0.9998071	0.9992956	0.9999472
		B (dorsal-oblique)	0.9996497	0.9988323	0.9998950
		C (front/back-oblique)	0.9998078	0.9993291	0.9999450
	<i>Panstrongylus</i>	A	0.9999982	0.9999683	0.9999999
		B	0.9999967	0.9999446	0.9999998
		C	0.9999982	0.9999690	0.9999999
AdaBoost	<i>Triatoma</i>	A	0.9970456	0.9941101	0.9985203
		B	0.9946466	0.9913515	0.9966904
		C	0.9970568	0.9947102	0.9983641
	<i>Panstrongylus</i>	A	0.9999718	0.9996091	0.9999980
		B	0.9999489	0.9993224	0.9999961
		C	0.9999719	0.9996198	0.9999979
Gradient Boosting	<i>Triatoma</i>	A	0.9936410	0.9882059	0.9965801
		B	0.9885091	0.9833282	0.9920929
		C	0.9936649	0.9895485	0.9961662
	<i>Panstrongylus</i>	A	0.9999392	0.9991707	0.9999955
		B	0.9998895	0.9985629	0.9999915
		C	0.9999394	0.9991935	0.9999955
Histogram Boosting	<i>Triatoma</i>	A	0.9967809	0.9937743	0.9983380
		B	0.9941682	0.9909802	0.9962337
		C	0.9967931	0.9944386	0.9981526
	<i>Panstrongylus</i>	A	0.9999693	0.9995774	0.9999978
		B	0.9999443	0.9992674	0.9999958
		C	0.9999694	0.9995890	0.9999977
Linear Discriminant	<i>Triatoma</i>	A	0.9968862	0.9939598	0.9983970
		B	0.9943583	0.9912346	0.9963729
		C	0.9968979	0.9946019	0.9982191
	<i>Panstrongylus</i>	A	0.9999703	0.9995909	0.9999978
		B	0.9999461	0.9992907	0.9999959
		C	0.9999704	0.9996021	0.9999978

Supplementary figures

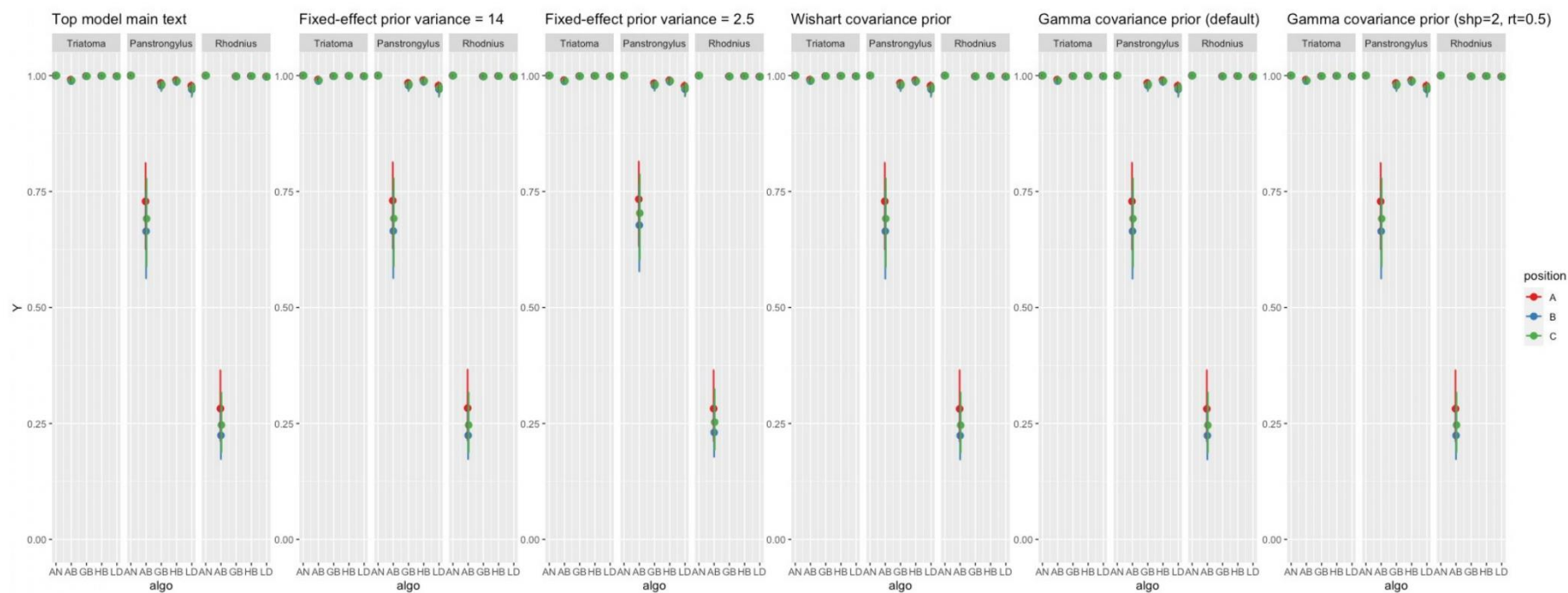


Fig. S1. Prior robustness check for accuracy. From left to right, predictions of the top-ranking model as presented in the main text and with the following alternative prior specifications: $N(0,14)$ fixed-effect priors; $N(0,2.5)$ fixed-effect priors; Wishart covariance priors, which *blme* uses by default; gamma(shape=2.5, rate=0), the default gamma specification in *blme*; and gamma(shape=2.5, rate=0.5).

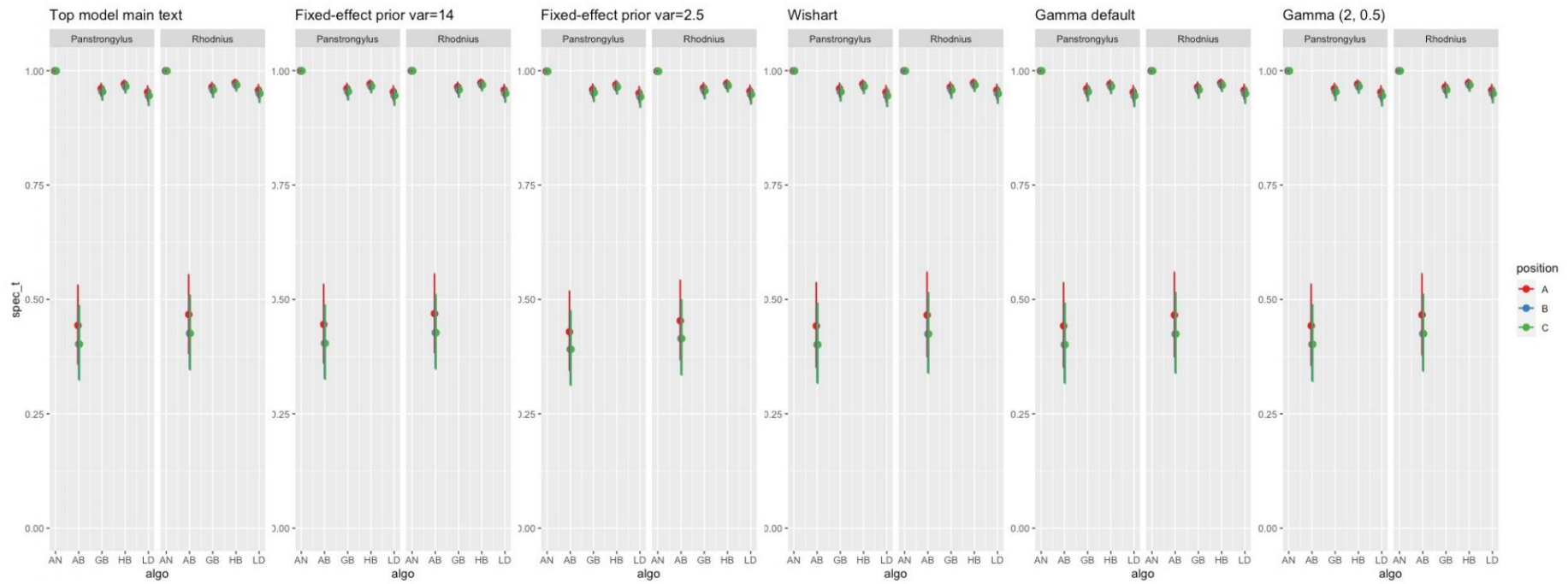


Fig. S2. Prior robustness check for specificity – *Triatoma*. From left to right, predictions of the top-ranking model as presented in the main text and with the following alternative prior specifications: $N(0,14)$ fixed-effect priors; $N(0,2.5)$ fixed-effect priors; Wishart covariance priors, which *blme* uses by default; gamma(shape=2.5, rate=0), the default gamma specification in *blme*; and gamma(shape=2.5, rate=0.5).

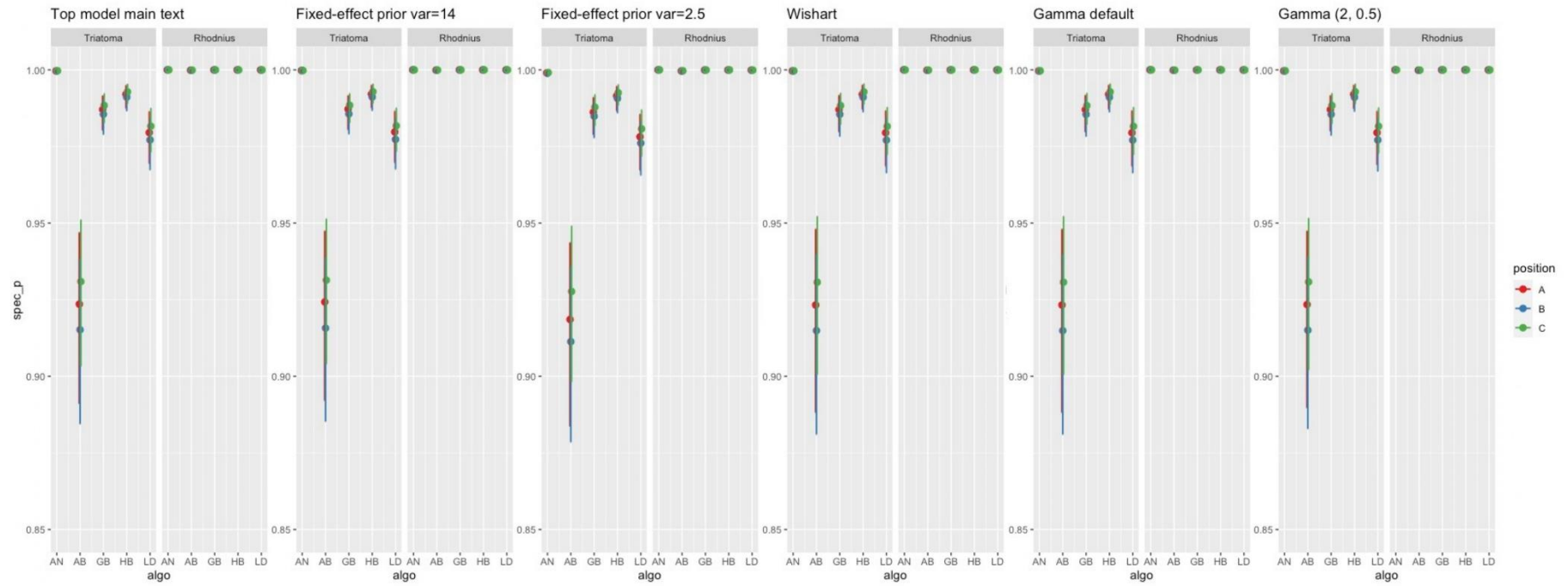


Fig. S3. Prior robustness check for specificity – *Panstrongylus*. From left to right, predictions of the top-ranking model as presented in the main text and with the following alternative prior specifications: $N(0,14)$ fixed-effect priors; $N(0,2.5)$ fixed-effect priors; Wishart covariance priors, which *blme* uses by default; $\text{gamma}(\text{shape}=2.5, \text{rate}=0)$, the default gamma specification in *blme*; and $\text{gamma}(\text{shape}=2.5, \text{rate}=0.5)$.

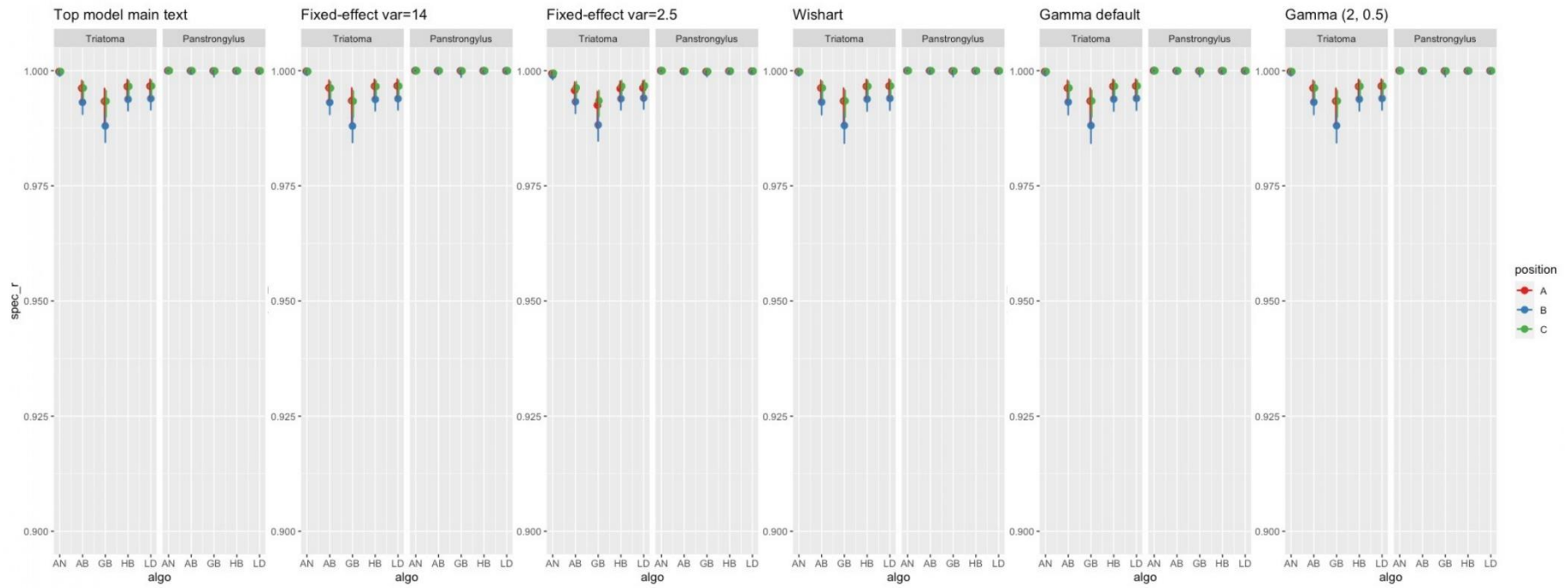


Fig. S4. Prior robustness check for specificity – *Rhodnius*. From left to right, predictions of the top-ranking model as presented in the main text and with the following alternative prior specifications: $N(0,14)$ fixed-effect priors; $N(0,2.5)$ fixed-effect priors; Wishart covariance priors, which *blme* uses by default; gamma(shape=2.5, rate=0), the default gamma specification in *blme*; and gamma(shape=2.5, rate=0.5)

CAPÍTULO II

Identificação automática de imagens dorsais de triatomíneos adultos ao nível de espécie usando a rede neural convolucional pré-treinada AlexNet

Manuscrito em elaboração

Resumo

Existem 160 espécies de triatomíneos descritas com diferentes capacidades de transmissão de *Trypanosoma cruzi*, agente etiológico da doença de Chagas. Apesar dos avanços para obtenção de um sistema acurado para identificação automática de triatomíneos, uma análise com um banco de imagens abundante e diverso é fundamental para o desenvolvimento de um sistema global de identificação desses vetores. Nosso objetivo foi avaliar o desempenho de uma rede de aprendizagem profunda (AlexNet) para identificação de 65 espécies de triatomíneos a partir de um banco de imagens dorsais de insetos adultos, com diversas qualidades e avaliar seu desempenho considerando somente as espécies que já foram reportadas em casas e infectadas por *T. cruzi*. Nós usamos uma amostra de 6397 triatomíneos pertencentes a 65 espécies de sete gêneros. AlexNet escolheu aleatoriamente o conjunto de dados de treinamento (70%) e o conjunto de dados para teste (20%). Os 10% restantes foram utilizados para a validação interna do algoritmo durante o treinamento, realizado no MATLAB. AlexNet apresentou uma acurácia de ~0,93 (IC95%, 0,913-0,941) para identificar 65 espécies de triatomíneos a partir de fotografias com diversas qualidades de resolução. A acurácia específica alcançou valor máximo para 21 espécies, principalmente dos gêneros *Rhodnius* e *Panstrongylus*. O desempenho foi melhorado para ~0,95 (IC95%, 0,930-0,958) quando foram consideradas apenas as espécies com maior capacidade vetorial. Esses resultados evidenciam que a rede AlexNet, quando alimentada com um banco de imagens de treinamento amplo, diversificado e bem estruturado, exibe um desempenho altamente promissor para identificação de espécies de triatomíneos. Nosso estudo fornece a base para o desenvolvimento de um sistema automatizado de identificação dos vetores da doença de Chagas com participação comunitária em larga escala, o que poderá contribuir para a ciência cidadã.

Palavras-Chave: Triatominae, deep learning, vigilância entomológica, ciência cidadã.

1. Introdução

Os triatomíneos (Hemiptera: Reduviidae) são vetores do parasito *Trypanosoma cruzi* Chagas, 1909, agente etiológico da doença de Chagas (Lent & Wygodzinsky, 1979; WHO, 2021). Atualmente, existem 160 espécies de triatomíneos descritas com diferentes capacidades vetoriais, classificadas em 16 gêneros e cinco

tribos (Monteiro *et al.*, 2018, Zhao *et al.*, 2023), dos quais a maioria pode ser identificada utilizando caracteres morfológicos com o auxílio de chaves dicotômicas convencionais ou ilustradas (Lent & Wygodzinsky, 1979; Galvão, 2014). Entretanto, essas chaves de identificação estão desatualizadas e seu uso é limitado, pois geralmente necessita de taxonomistas experientes e demanda muito tempo de execução (Carvalho *et al.* 2007). Uma alternativa é o uso de chaves eletrônicas, que têm facilitado a identificação de triatomíneos, como o TriatoKey (Oliveira *et al.*, 2017) e o TriatoDex (Gurgel-Gonçalves *et al.*, 2021), porém, apesar de um destes ter a vantagem de possuir um especialista para solucionar dúvidas (TriatoKey), na maior parte de suas execuções ainda necessitam de usuários treinados para diferenciar as características morfológicas. Nesse contexto, a utilização de técnicas de aprendizado de máquina pode ser uma aplicação promissora para auxiliar o serviço de identificação de triatomíneos, de forma automática.

Métodos para identificação automática de insetos têm obtido bons resultados nos últimos anos (Valan, *et al.*, 2019; Høye *et al.*, 2021; Ong *et al.*, 2022; Kittichai *et al.*, 2023). Para triatomíneos, a aplicação de métodos de aprendizado de máquina para identificação automática de imagens vem sendo aplicados (Gurgel-Gonçalves *et al.*, 2017; Khalighifar *et al.*, 2019; Cruz *et al.*, 2020; Abdelghani *et al.*, 2021; Cochero *et al.*, 2022). Um sistema de identificação visual automática para 39 espécies de triatomíneos do Brasil e 12 espécies do México foi desenvolvido a partir de fotografias de alta qualidade e processamento das imagens, resultando em sucesso de identificação superior a 80% (Gurgel-Gonçalves *et al.*, 2017). Posteriormente, em um estudo utilizando os mesmos conjuntos de fotografias, e aplicando um algoritmo baseado em deep learning, foi alcançado até 98,9% de identificações corretas (Khalighifar *et al.*, 2019). Neste contexto, um passo importante para vigilância com participação comunitária seria identificar triatomíneos de diversas espécies usando imagens de diferentes qualidades.

Apesar dos esforços para obtenção de um sistema treinado e acurado para identificação de triatomíneos (Gurgel-Gonçalves *et al.*, 2017; Khalighifar *et al.*, 2019; Parsons & Banitaan, 2021; Abdelghani *et al.*, 2021; Cochero *et al.*, 2022), uma análise com um banco de imagens abundante e diverso é fundamental para o desenvolvimento de um sistema global de identificação desses vetores. Cuidados devem ser necessários durante a construção de bancos de imagens, e conseqüentemente, durante as avaliações do aprendizado de máquina, para não

haver identificações incorretas, mesmo que o modelo de identificação apresente desempenho promissor (e.g. Parsons & Banitaan, 2021), e deve ser observadas as incertezas em volta das probabilidades de acertos (acurácia) preditas pelos algoritmos avaliados (ver capítulo anterior) e verificado o progresso de treinamento para estabelecer os parâmetros que melhor se adaptam ao banco de imagens adotado (treino, teste, validação interna, número de épocas, etc).

Neste estudo, aplicamos uma rede neural convolucional pré-treinada chamada AlexNet (Krizhevsky *et al.*, 2012), para identificar espécies de triatomíneos. Esse algoritmo tem sido usado com frequência em tarefas para identificação da diversidade biológica, inclusive para vetores de doenças tropicais, como espécies do gênero *Aedes* (e.g. Raitoharju *et al.*, 2016; Marques *et al.*, 2018; Motta *et al.*, 2019; Likith *et al.*, 2021) e triatomíneos (Miranda *et al.* 2023, capítulo 1). Dessa forma, considerando a importância médica dos triatomíneos como vetores da doença de Chagas e a necessidade de se obter uma identificação rápida e correta das espécies desse grupo, buscamos: 1) avaliar o desempenho da rede AlexNet na identificação de 65 espécies de triatomíneos a partir de um banco de imagens com diversas qualidades; e 2) avaliar o desempenho do AlexNet quando consideramos apenas 48 espécies do banco de imagens, nas quais já foram reportadas invadindo residências e naturalmente infectadas por *T. cruzi*. Nossas hipóteses sugerem que para estas tarefas o AlexNet terá uma probabilidade de acerto superior a 90%, que poderá melhorar quando for considerado menos espécies nas análises e que existe uma correlação positiva entre o número de fotografias por espécie e o percentual de identificações corretas.

2. Material e métodos

2.1. Banco de imagens dos insetos

Nós usamos uma amostra de 6.397 triatomíneos pertencentes a 65 espécies de sete gêneros: *Cavernicola* Barber, 1937 (n = 76 espécimes), *Dipetalogaster* Usinger, 1939 (n = 54), *Eratyrus* Stål, 1859 (n = 106) *Panstrongylus* Berg, 1879 (n = 1.093), *Psammolestes* Bergroth, 1911 (n = 307), *Rhodnius* Stål, 1859 (n = 1.375) e *Triatoma* Laporte, 1832 (n = 3.386) (Apêndice 1, Tabela 1). A identificação dos exemplares foi confirmada utilizando chaves impressas (Lent & Wygodzinsky, 1979; Galvão *et al.*, 2014) e a chave eletrônica TRIATODEX (Gurgel-Gonçalves *et al.*, 2021),

além de informações sobre a origem geográfica dos espécimes. Assim nós reunimos o banco de imagens disponibilizado em Gurgel-Gonçalves *et al.* (2017) e juntamos a um novo conjunto de imagens obtidas de espécimes disponíveis no laboratório de Parasitologia Médica e Biologia de Vetores da Universidade de Brasília (UnB), no Laboratório de Parasitologia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) e no Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos do Instituto Oswaldo Cruz (Fiocruz-RJ), usando câmeras de celulares (marca Motorola - modelos moto G6 play e moto G22) , formando um grande banco de imagens dorsais de baixa e alta qualidade de resolução e variação na cor de fundo.

2.2. Algoritmo: treino e teste

Nós usamos AlexNet, um algoritmo de aprendizagem profunda que implementa uma Rede Neural Convolutiva (CNN) que realiza o aprendizado de transferência, tomando uma rede pré-treinada como fonte de alguns recursos de entrada antes de aprender com as imagens do conjunto de dados. As CNNs são um tipo especializado de rede neural para processar imagens de entrada que emprega uma convolução linear (Goodfellow *et al.*, 2016). O algoritmo de rede pré-treinado usado é uma CNN treinada com um grande conjunto de imagens (1,2 milhão de imagens de alta resolução, segmentadas em 1000 classes) (Krizhevsky *et al.*, 2012). As imagens originais foram redimensionadas para obter uma representação de 227×227 pixels, usada como entrada da rede AlexNet. No procedimento de aprendizado, o AlexNet escolheu aleatoriamente o conjunto de dados de treinamento (70%) e o conjunto de dados para teste (20%). Os 10% restantes foram utilizados para a validação interna do algoritmo durante o treinamento. Ao dividir os dados dessa maneira, a avaliação do desempenho do modelo se torna independente do conjunto de treinamento, o que é essencial para garantir que a rede neural seja capaz de generalizar bem para novos exemplos. Para o treinamento, utilizamos o otimizador "stochastic gradient descent with momentum" com as configurações padrão da rede, exceto para a taxa de aprendizado inicial (que definimos como 0.001) e o número máximo de épocas (que definimos como 100 após testes preliminares para observar como o desempenho da rede se comportava). Estes procedimentos foram realizados no ambiente do software computacional MATLAB (www.mathworks.com). Resultados binários das identificações (corretas/incorretas) foram usados como variáveis de nas

análises de acordo com (Miranda *et al.* 2023, capítulo 1). No apêndice 2 apresentamos um protocolo de uso da rede AlexNet aplicada aos dados do nosso estudo.

2.3. Análise de dados

Nós realizamos dois experimentos com o AlexNet. No primeiro, treinamos o algoritmo para aprender a identificar 65 espécies de triatomíneos sendo utilizadas 1285 imagens de teste. No segundo, incluímos apenas as espécies que, dentre as que consideramos neste estudo, já haviam sido reportadas em casas e infectadas por *T. cruzi* segundo Abad-Franch & Gurgel-Gonçalves (2021) totalizando 1029 imagens de teste. Analisamos as curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda de validação para destacar tendências gerais do treinamento. Após o treino do algoritmo, nós avaliamos o seu desempenho com a quantidade de erros de identificação e os resultados de acurácia da rede (probabilidade de acerto). A acurácia foi calculada da seguinte forma:

$$\text{Acurácia geral} = \frac{VP+VN}{VP+VN+FP+FN}$$

A sensibilidade, que pode ser interpretada como uma acurácia específica, ou seja, uma probabilidade de acerto para cada espécie, também foi calculada:

$$\text{Acurácia específica} = \frac{VP}{VP+FN}$$

VP = Verdadeiros positivos

VN = Verdadeiros negativos

FP = Falsos positivos

FN = Falsos negativos

Nós apresentamos nossos resultados em tabelas descritivas e gráficos ilustrados e realizamos análises exploratórias simples calculando as frequências e proporções com intervalos de confiança (ICs) de 'Wilson' de 95% (Newcombe, 1998) usando o pacote R 'Hmisc' (Harrell e Dupont, 2023) e para facilitar a interpretação dos resultados, separamos as ilustrações por grupos seguindo a sequência: 'espécies do gênero *Triatoma*'; 'espécies do gênero *Panstrongylus*'; 'espécies do gênero *Rhodnius*'; e 'espécies dos demais gêneros'. Analisamos os dados usando o software computacional R 4.2.1 (R Core Team, 2022) e a interface RStudio 2023.03.1.446 (Posit Software, 2022).

3. Resultados

3.1. Resultado do treino

Os gráficos abaixo (Figura 1) permitem acompanhar o progresso do treinamento da rede ao longo das épocas e iterações. À medida que as épocas progredem, a precisão nos dados de treinamento e validação aumenta e as perdas diminuem, o que indica que a rede está aprendendo bem com os novos dados. É importante notar que a precisão de validação é usada para avaliar a acurácia da rede com dados não vistos durante o treinamento, o que é um bom indicador do desempenho geral da rede. O treinamento é concluído após 100 épocas, como especificado nos parâmetros de treinamento.

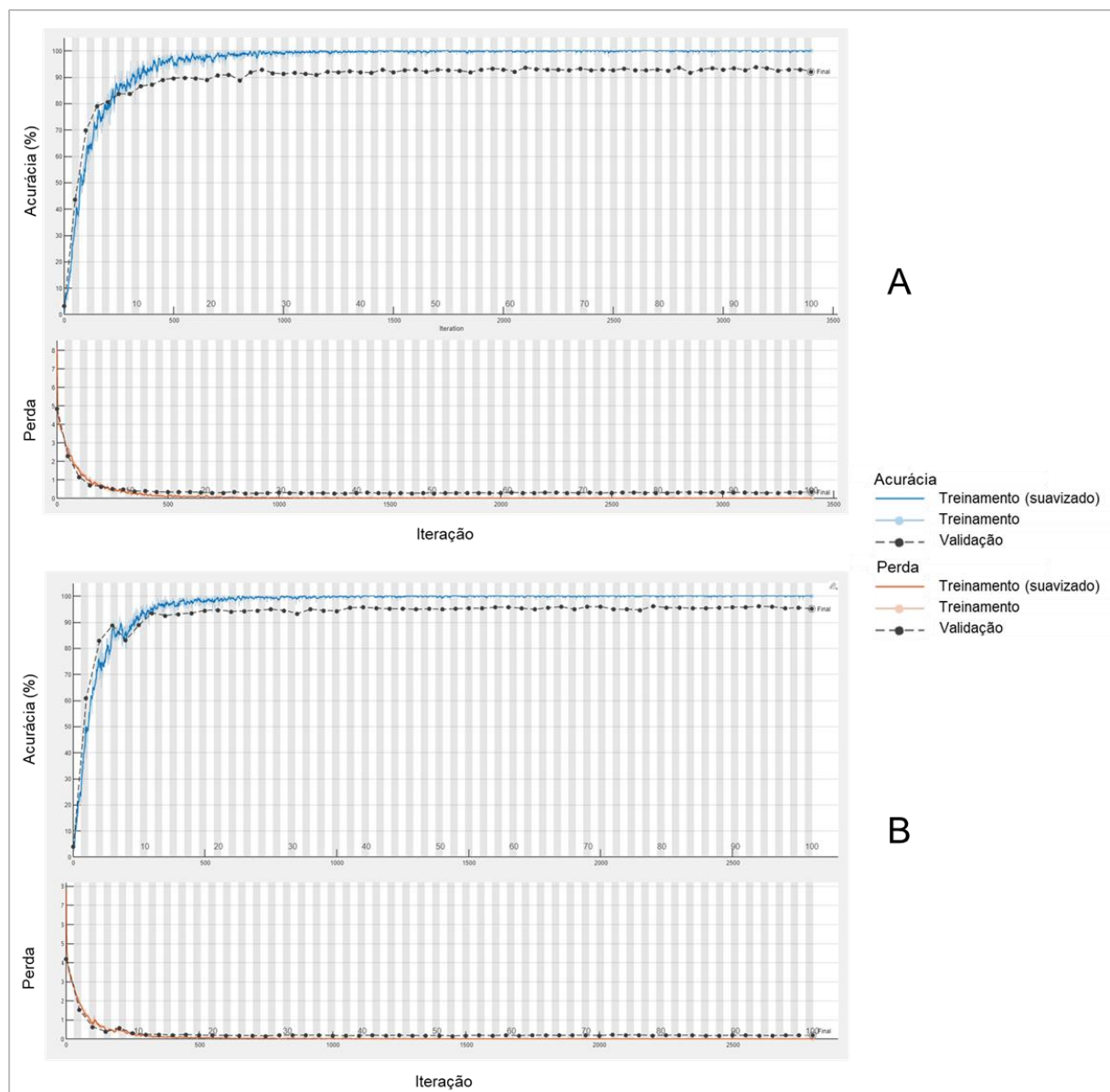


Figura 1. Progresso de treinamento da rede AlexNet. A – Treino incluindo 65 espécies com acurácia de validação de 92,13%. B – Treino incluindo 48 espécies com acurácia de validação

de 95,07%. As métricas de acurácia e perda são exibidas ao longo das iterações. As curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda de validação são apresentadas, com uma suavização aplicada para destacar tendências gerais. Os eixos horizontal e vertical representam as iterações e os valores correspondentes das métricas, respectivamente.

3.2. Desempenho da AlexNet para identificação de 65 espécies de triatomíneos

A acurácia geral da rede AlexNet para identificação das 65 espécies de triatomíneos foi de ~0,928 (IC95%, 0,913-0,941). A acurácia específica alcançou valor máximo para 21 espécies (Tabela 1), enquanto as acurácias abaixo de 0,800 variaram de ~0,778 a ~0,333. Acurácias específicas menores mostraram que os principais erros de identificação ocorreram com *T. carcavalloei*, com duas imagens sendo confundidas com *T. pintodiasi* e uma imagem confundida com *T. rubrovaria*; *T. juazeirensis*, com 3 imagens confundidas com *T. melanica*; *T. petrocchiaei*, com três imagens confundidas com *T. brasiliensis*, *T. vandaei* e *T. melanica*. *T. matogrossensis*, com três imagens confundidas com *T. vandaei* e outras três confundidas com *T. maculata*, *T. sordida*, *P. lignarius*. Três imagens de *T. phyllosoma* foram confundidas com *T. mazzotti* e uma com *T. pallidipennis*. As demais informações sobre os erros de identificação podem ser consultadas no apêndice 3, Tabela 2.

Tabela 1. Acurácia da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a 65 espécies de triatomíneos com intervalo de confiança de 95%. As espécies com maior capacidade vetorial foram destacadas em negrito.

Espécie	"N"	Acurácia			
		Correto	Proporção	IC inferior	IC superior
<i>Cavernicola lenti</i> Barrett and Arias, 1985	12	12	1,000	0,758	1,000
<i>Cavernicola pilosa</i> Barber, 1937	3	2	0,667	0,208	0,939
<i>Dipetalogaster maxima</i> (Uhler, 1894)	11	11	1,000	0,741	1,000
<i>Eratyrus cuspidatus</i> Stål, 1859	3	3	1,000	0,439	1,000
<i>Eratyrus mucronatus</i> Stål, 1859	19	18	0,947	0,754	0,991
<i>Panstrongylus diasi</i> Pinto & Lent, 1946	18	18	1,000	0,824	1,000
<i>Panstrongylus geniculatus</i> (Latreille, 1811)	25	25	1,000	0,867	1,000
<i>Panstrongylus lignarius</i> (Walker, 1873)	70	69	0,986	0,923	0,997
<i>Panstrongylus lutzi</i> (Neiva & Pinto, 1923)	19	19	1,000	0,832	1,000
<i>Panstrongylus megistus</i> (Burmeister, 1835)	45	45	1,000	0,921	1,000
<i>Panstrongylus rufotuberculatus</i> (Champion, 1899)	5	3	0,600	0,231	0,882
<i>Panstrongylus tibiamaculatus</i> (Pinto, 1926)	36	36	1,000	0,904	1,000
<i>Psammolestes coreodes</i> Bergroth, 1911	27	21	0,778	0,592	0,894
<i>Psammolestes tertius</i> Lent & Jurberg, 1966	34	31	0,912	0,770	0,970
<i>Rhodnius brethesi</i> Matta, 1919	31	31	1,000	0,890	1,000
<i>Rhodnius domesticus</i> Neiva & Pinto, 1923	15	14	0,933	0,702	0,988
<i>Rhodnius marabaensis</i> Souza et al., 2016	11	10	0,909	0,623	0,984

<i>Rhodnius milesi</i> Carcavallo <i>et al.</i> , 2001	11	11	1,000	0,741	1,000
<i>Rhodnius montenegrensis</i> Rosa <i>et al.</i> , 2012	20	19	0,950	0,764	0,991
<i>Rhodnius nasutus</i> Stål, 1859	51	48	0,941	0,841	0,980
<i>Rhodnius neglectus</i> Lent, 1954	53	53	1,000	0,932	1,000
<i>Rhodnius pictipes</i> Stål, 1872	35	34	0,971	0,855	0,995
<i>Rhodnius prolixus</i> Stål, 1859	15	14	0,933	0,702	0,988
<i>Rhodnius robustus</i> Larrousse, 1927	7	6	0,857	0,487	0,974
<i>Rhodnius stali</i> Jurberg & Galvão, 1993	27	26	0,963	0,817	0,993
<i>Triatoma arthurneivai</i> Lent & Martins, 1940	10	8	0,800	0,490	0,943
<i>Triatoma baratai</i> Carcavallo & Jurberg, 2000	13	13	1,000	0,772	1,000
<i>Triatoma barberi</i> Usinger, 1939	6	6	1,000	0,610	1,000
<i>Triatoma brasiliensis</i> Neiva, 1911	52	48	0,923	0,818	0,970
<i>Triatoma breyeri</i> Del Ponte, 1929	7	6	0,857	0,487	0,974
<i>Triatoma carcavalloei</i> Jurberg <i>et al.</i> , 1998	11	6	0,545	0,280	0,787
<i>Triatoma circummaculata</i> (Stål, 1859)	5	5	1,000	0,566	1,000
<i>Triatoma costalimai</i> Verano & Galvão, 1958	38	34	0,895	0,759	0,958
<i>Triatoma delpontei</i> Romaña & Abalos, 1947	10	9	0,900	0,596	0,982
<i>Triatoma dimidiata</i> (Latreille, 1811)	34	33	0,971	0,851	0,995
<i>Triatoma eratyrsiformis</i> Del Ponte, 1929	3	2	0,667	0,208	0,939
<i>Triatoma gerstaeckeri</i> (Stål, 1859)	3	1	0,333	0,061	0,792
<i>Triatoma guazu</i> Lent & Wygodzinsky, 1979	7	4	0,571	0,250	0,842
<i>Triatoma infestans</i> (Klug, 1834)	49	45	0,918	0,808	0,968
<i>Triatoma juazeirensis</i> Costa & Felix, 2007	7	4	0,571	0,250	0,842
<i>Triatoma jurbergi</i> Carcavallo <i>et al.</i> , 1998	8	8	1,000	0,676	1,000
<i>Triatoma klugi</i> Carcavallo <i>et al.</i> , 2001	8	7	0,875	0,529	0,978
<i>Triatoma lenti</i> Sherlock & Serafim, 1967	19	16	0,842	0,624	0,945
<i>Triatoma longipennis</i> Usinger, 1939	11	10	0,909	0,623	0,984
<i>Triatoma maculata</i> (Erichson, 1848)	30	28	0,933	0,787	0,982
<i>Triatoma matogrossensis</i> Leite & Barbosa, 1953	21	15	0,714	0,500	0,862
<i>Triatoma mazzottii</i> Usinger, 1941	5	4	0,800	0,376	0,964
<i>Triatoma melanica</i> Neiva & Lent, 1941	23	22	0,957	0,790	0,992
<i>Triatoma melanocephala</i> Neiva & Pinto, 1923	21	21	1,000	0,845	1,000
<i>Triatoma mexicana</i> (Herrich-Schaeffer, 1848)	9	9	1,000	0,701	1,000
<i>Triatoma nítida</i> Usinger, 1939	3	3	1,000	0,439	1,000
<i>Triatoma pallidipennis</i> (Stål, 1872)	10	10	1,000	0,722	1,000
<i>Triatoma petrocchiai</i> Pinto & Barreto, 1925	9	6	0,667	0,354	0,879
<i>Triatoma phyllosoma</i> (Burmeister, 1835)	12	8	0,667	0,391	0,862
<i>Triatoma pintodiasi</i> Jurberg, <i>et al.</i> , 2013	15	14	0,933	0,702	0,988
<i>Triatoma platenses</i> Neiva, 1913	8	5	0,625	0,306	0,863
<i>Triatoma pseudomaculata</i> Correa & Espinola, 1964	32	30	0,938	0,799	0,983
<i>Triatoma rubrofaciata</i> (De Geer, 1773)	10	10	1,000	0,722	1,000
<i>Triatoma rubrovaria</i> (Blanchard, 1843)	55	54	0,982	0,904	0,997
<i>Triatoma sherlocki</i> Papa <i>et al.</i> , 2002	20	18	0,900	0,699	0,972
<i>Triatoma sordida</i> (Stål, 1859)	45	44	0,978	0,884	0,996
<i>Triatoma vandae</i> Carcavallo <i>et al.</i> , 2002	15	15	1,000	0,796	1,000
<i>Triatoma vitticeps</i> (Stål, 1859)	25	22	0,880	0,700	0,958
<i>Triatoma williami</i> Galvão <i>et al.</i> , 1965	7	6	0,857	0,487	0,974
<i>Triatoma wygodzinskyi</i> Lent, 1951	6	5	0,833	0,436	0,970
Acurácia geral	1285	1193	0,928	0,913	0,941

A acurácia específica para cada gênero apresentou uma variação maior dentro do gênero *Triatoma*, que representou o maior número de espécies estudadas.

Apenas 10 das 40 espécies de *Triatoma* obtiveram 1,000 de acurácia específica, enquanto a menor acurácia específica foi do *T. gerstaeckeri*, com ~0,333 (Figura 2).

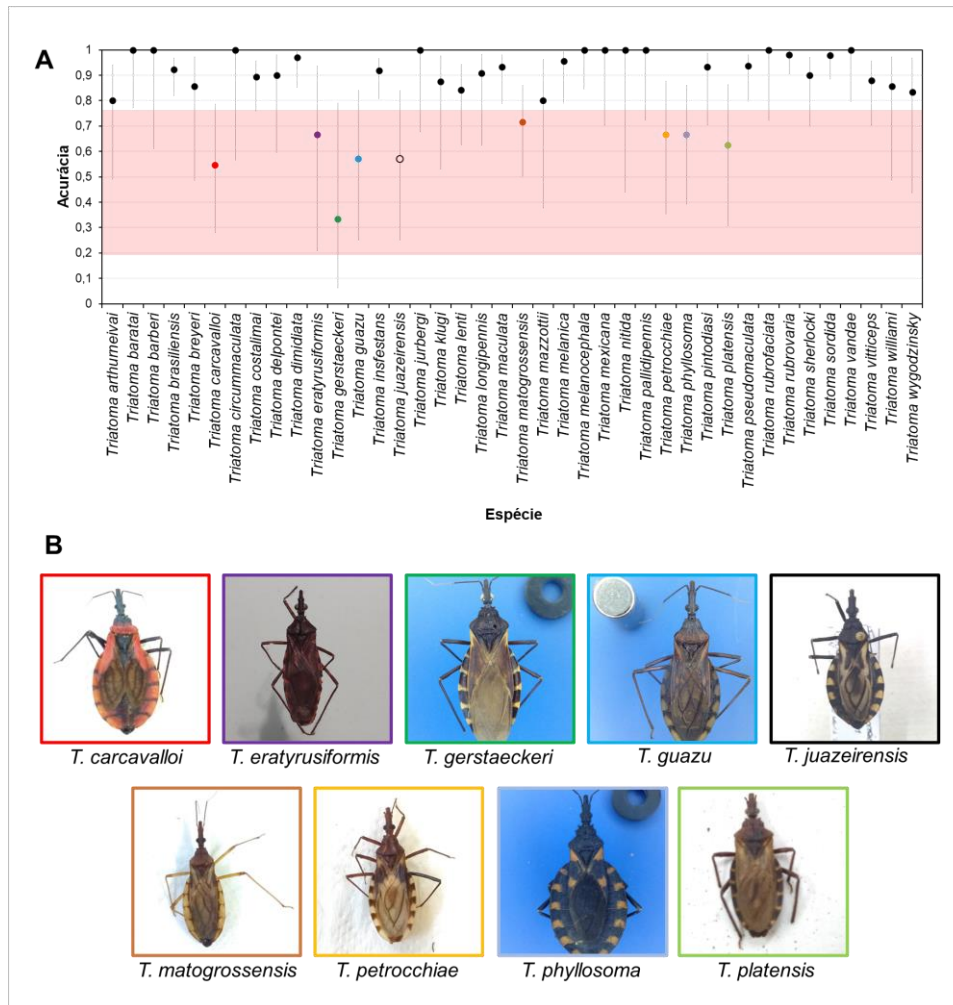


Figura 2. A – Acurácia específica da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a espécies do gênero *Triatoma* com intervalos de confiança de Wilson de 95%. Os pontos coloridos representam as espécies com valores de acurácia inferiores a 0,800. B – Exemplos das espécies que obtiveram acurácias específicas inferiores a 0,800. As imagens seguem a mesma sequência de distribuição dos pontos no gráfico.

A acurácia específica para as sete espécies do gênero *Panstrongylus* foi de 1,000 para *P. diasii*, *P. geniculatus*, *P. lutzi*, *P. megistus* e *P. tibiamaculatus*. Já *P. lignarius* obteve ~0,986 e *P. rufotuberculatus* ~0,600 (Figura 2).

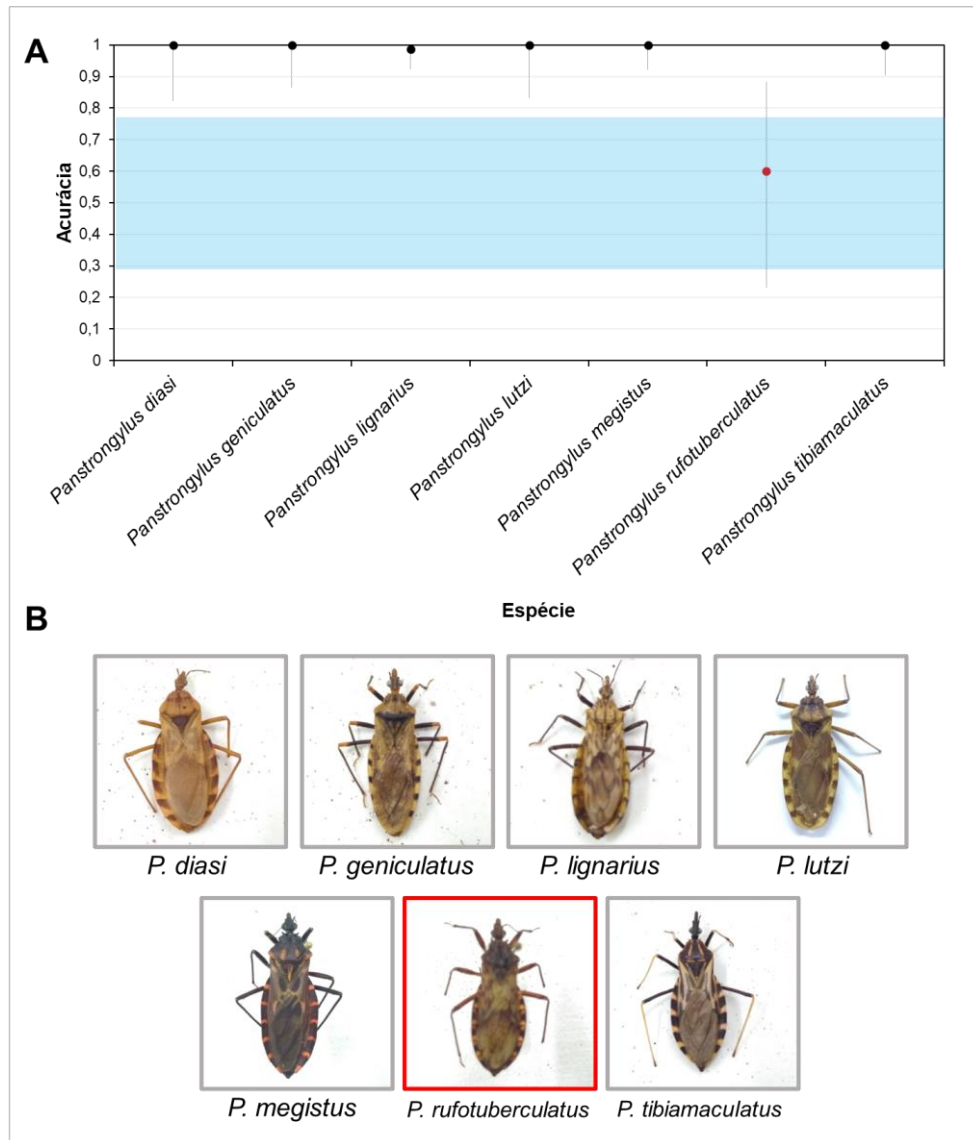


Figura 3. A – Acurácia específica da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a espécies do gênero *Panstrongylus* com intervalos de confiança de Wilson de 95%. Os pontos coloridos representam as espécies com valores de acurácia inferiores a 0,800. B – Exemplos das espécies do gênero *Panstrongylus* que foram incluídas. As imagens seguem a mesma sequência de distribuição dos pontos no gráfico.

A acurácia específica para as 11 espécies do gênero *Rhodnius* foi de 1,000 para *R. brethesi*, *R. milesi* e *R. neglectus*. Para as demais espécies do gênero a acurácia variou de ~0,857 a ~0,971, para *R. robustus* e *R. pictipes*, respectivamente (Figura 3).

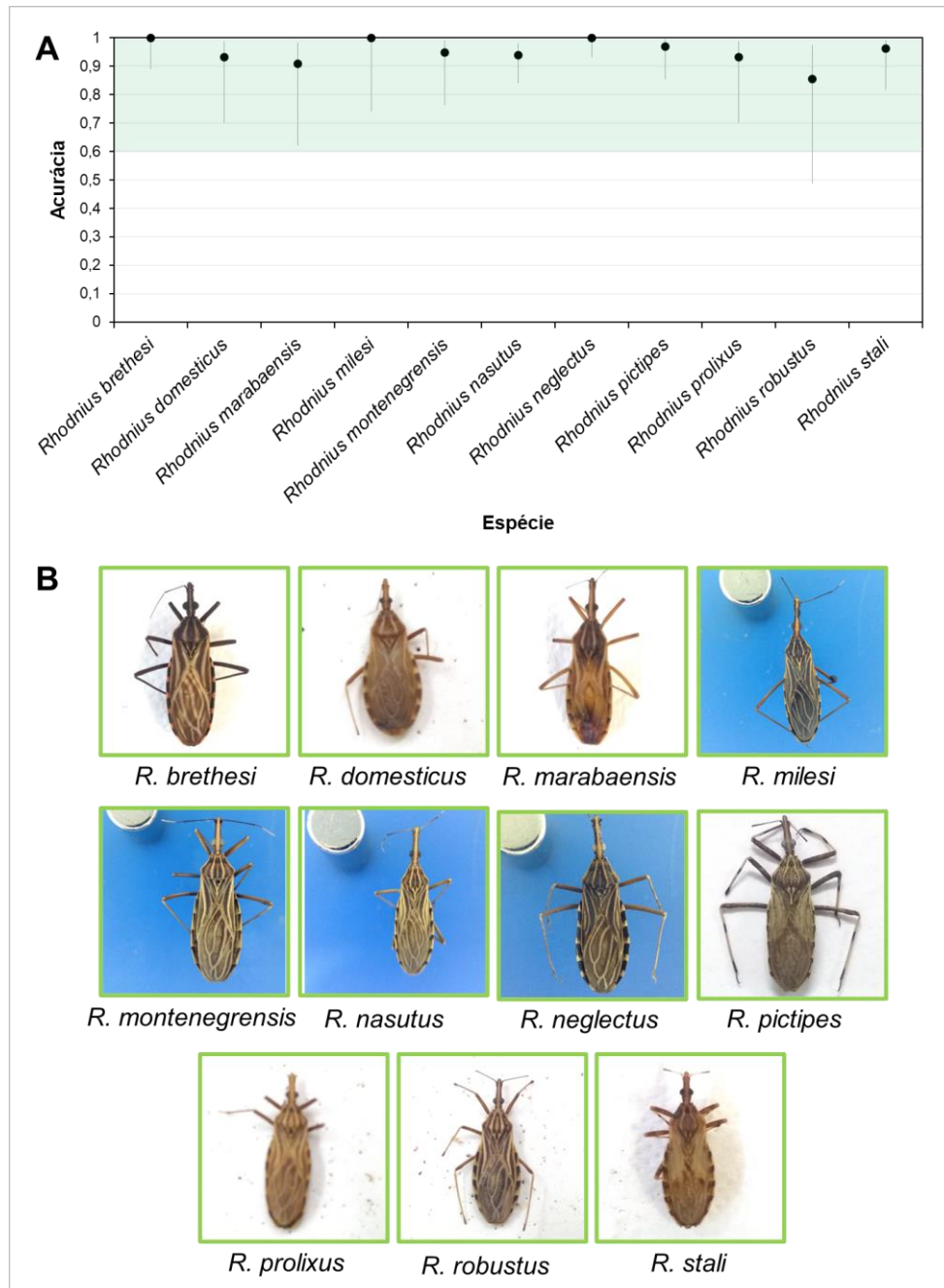


Figura 4. A – Acurácia específica da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a espécies do gênero *Rhodnius* com intervalos de confiança de Wilson de 95%. B – Exemplos das espécies do gênero *Rhodnius* que foram incluídas. As imagens seguem a mesma sequência de distribuição dos pontos no gráfico.

A acurácia específica para as espécies dos demais gêneros (*Cavernicola*, *Dipetalogaster*, *Eratyrus* e *Psammolestes*) foi de 1,000 para *C. lenti*, *D. maxima* e *E. cuspidatus*. Para as demais a acurácia variou de ~0,667 a ~0,947, para *C. pilosa* e *E. mucronatus*, respectivamente (Figura 4).

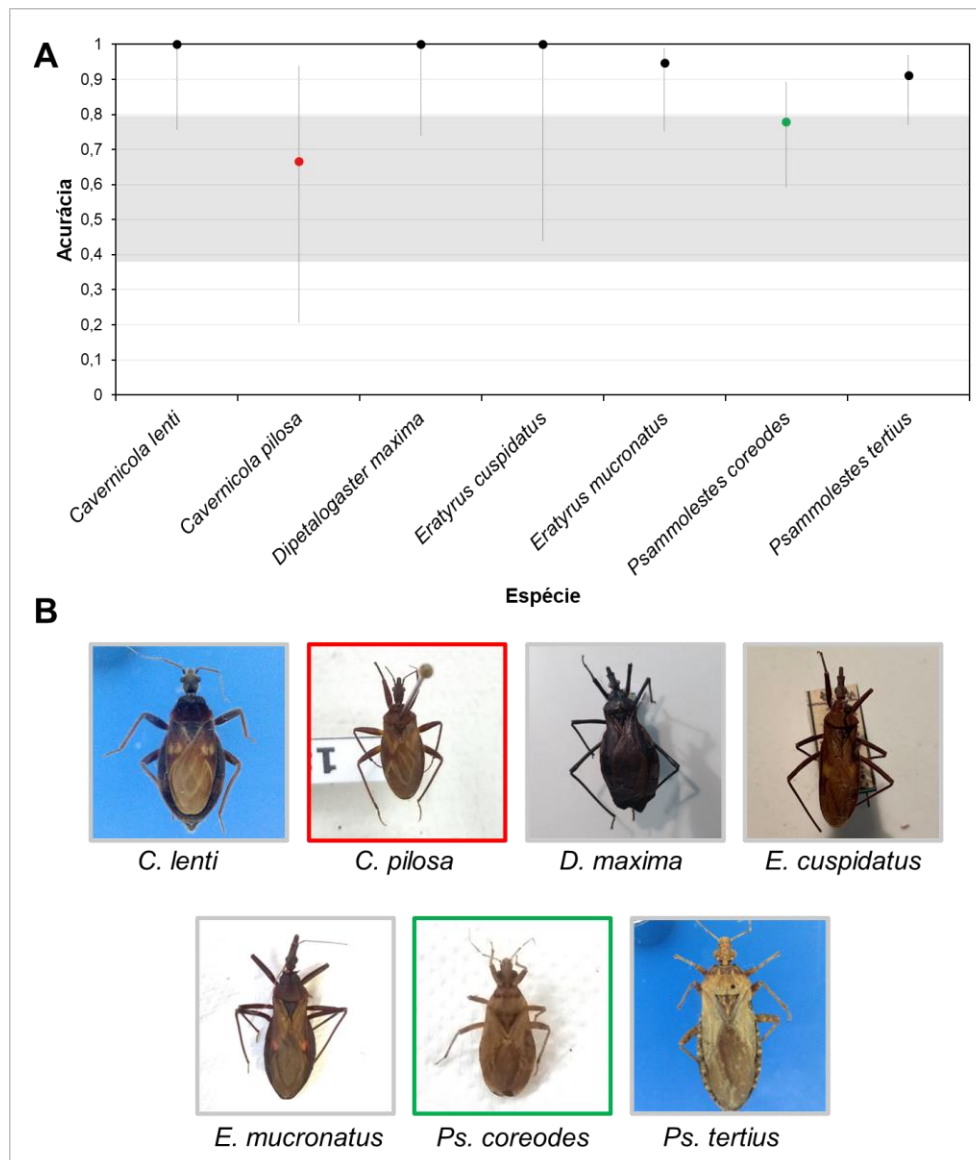


Figura 5. A – Acurácia específica da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a espécies dos gêneros *Cavernicola*, *Dipetalogaster*, *Eratyrus* e *Psammolestes* com intervalos de confiança de Wilson de 95%. Os pontos coloridos representam as espécies com valores de acurácia inferiores a 0,800. B – Exemplos das espécies que obtiveram acurácias específicas inferiores a 0,800. As imagens seguem a mesma sequência de distribuição dos pontos no gráfico.

As espécies com maiores registros de confusões na identificação foram *Ps. coreodes* (com 6 espécimes identificados como *Ps. tertius*), *T. carcavalloi* (com 1 espécime identificado como *T. arthurneivai*, 2 como *T. pintodiasi* e 2 como *T. rubrovaria*) e *T. matogrossensis* (com 1 espécime confundido com *T. maculata*, 1 com *T. sordida*, 3 com *T. vandae* e 1 com *P. lignarius*). Os demais erros de identificação pelo AlexNet são apresentados no Apêndice 3, Tabela 2.

3.3. Desempenho da AlexNet para identificação de 48 espécies de triatomíneos

A acurácia geral da rede AlexNet para identificação das 48 espécies de triatomíneos foi de ~0,946 (IC95%, 0,930-0,958). A acurácia específica alcançou valor máximo para 20 espécies (Tabela 2), enquanto as acurácias abaixo de 0,800 variaram de ~0,714 a ~0,600.

Tabela 2. Tabela 1. Acurácia da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a 48 espécies de triatomíneos com intervalo de confiança de 95%. As espécies com maior capacidade vetorial foram destacadas em negrito.

Espécie	"N"	Acurácia			
		Correto	Proporção	IC inferior	IC superior
<i>Dipetalogaster maxima</i>	11	10	0,909	0,623	0,984
<i>Eratyrus cuspidatus</i>	3	2	0,667	0,208	0,939
<i>Eratyrus mucronatus</i>	19	16	0,842	0,624	0,945
<i>Panstrongylus diasi</i>	18	18	1,000	0,824	1,000
<i>Panstrongylus geniculatus</i>	25	24	0,960	0,805	0,993
<i>Panstrongylus lignarius</i>	70	69	0,986	0,923	0,997
<i>Panstrongylus lutzi</i>	19	18	0,947	0,754	0,991
<i>Panstrongylus megistus</i>	45	45	1,000	0,921	1,000
<i>Panstrongylus rufotuberculatus</i>	5	3	0,600	0,231	0,882
<i>Panstrongylus tibiamaculatus</i>	36	36	1,000	0,904	1,000
<i>Rhodnius domesticus</i>	15	15	1,000	0,796	1,000
<i>Rhodnius montenegrensis</i>	20	20	1,000	0,839	1,000
<i>Rhodnius nasutus</i>	51	49	0,961	0,868	0,989
<i>Rhodnius neglectus</i>	53	52	0,981	0,901	0,997
<i>Rhodnius pictipes</i>	35	34	0,971	0,855	0,995
<i>Rhodnius prolixus</i>	15	13	0,867	0,621	0,963
<i>Rhodnius robustus</i>	7	6	0,857	0,487	0,974
<i>Rhodnius stali</i>	27	25	0,926	0,766	0,979
<i>Triatoma barberi</i>	6	6	1,000	0,610	1,000
<i>Triatoma brasiliensis</i>	52	48	0,923	0,818	0,970
<i>Triatoma breyeri</i>	7	7	1,000	0,646	1,000
<i>Triatoma carcavalloii</i>	11	11	1,000	0,741	1,000
<i>Triatoma circummaculata</i>	5	5	1,000	0,566	1,000
<i>Triatoma costalimai</i>	38	36	0,947	0,827	0,985
<i>Triatoma delpontei</i>	10	10	1,000	0,722	1,000
<i>Triatoma dimidiata</i>	34	34	1,000	0,898	1,000
<i>Triatoma eratyrusiformis</i>	3	3	1,000	0,439	1,000
<i>Triatoma gerstaeckeri</i>	3	3	1,000	0,439	1,000
<i>Triatoma infestans</i>	49	43	0,878	0,758	0,943
<i>Triatoma juazeirensis</i>	7	5	0,714	0,359	0,918
<i>Triatoma jurbergi</i>	8	8	1,000	0,676	1,000
<i>Triatoma lenti</i>	19	17	0,895	0,686	0,971
<i>Triatoma longipennis</i>	11	10	0,909	0,623	0,984
<i>Triatoma maculata</i>	30	26	0,867	0,703	0,947
<i>Triatoma mazzotti</i>	5	5	1,000	0,566	1,000
<i>Triatoma melanica</i>	23	20	0,870	0,679	0,955
<i>Triatoma melanocephala</i>	21	19	0,905	0,711	0,973
<i>Triatoma mexicana</i>	9	9	1,000	0,701	1,000

<i>Triatoma nitida</i>	3	3	1,000	0,439	1,000
<i>Triatoma pallidipennis</i>	10	10	1,000	0,722	1,000
<i>Triatoma petrocchiai</i>	9	8	0,889	0,565	0,980
<i>Triatoma phyllosoma</i>	12	10	0,833	0,552	0,953
<i>Triatoma platensis</i>	8	8	1,000	0,676	1,000
<i>Triatoma pseudomaculata</i>	32	29	0,906	0,758	0,968
<i>Triatoma rubrofasciata</i>	10	8	0,800	0,490	0,943
<i>Triatoma rubrovaria</i>	55	54	0,982	0,904	0,997
<i>Triatoma sherlocki</i>	20	18	0,900	0,699	0,972
<i>Triatoma sordida</i>	45	45	1,000	0,921	1,000
Total	1029	973	0,946	0,930	0,958

Nesta análise, das 30 espécies de *Triatoma* incluídas, 15 obtiveram acurácias de 1,000, enquanto as demais variaram de ~0,833 (IC 95%, 0,552-0,953) a ~0,947 (IC 95%, 0,827-0,985), para *Triatoma phyllosoma* e *T. costalimai*, respectivamente (Figura 6).

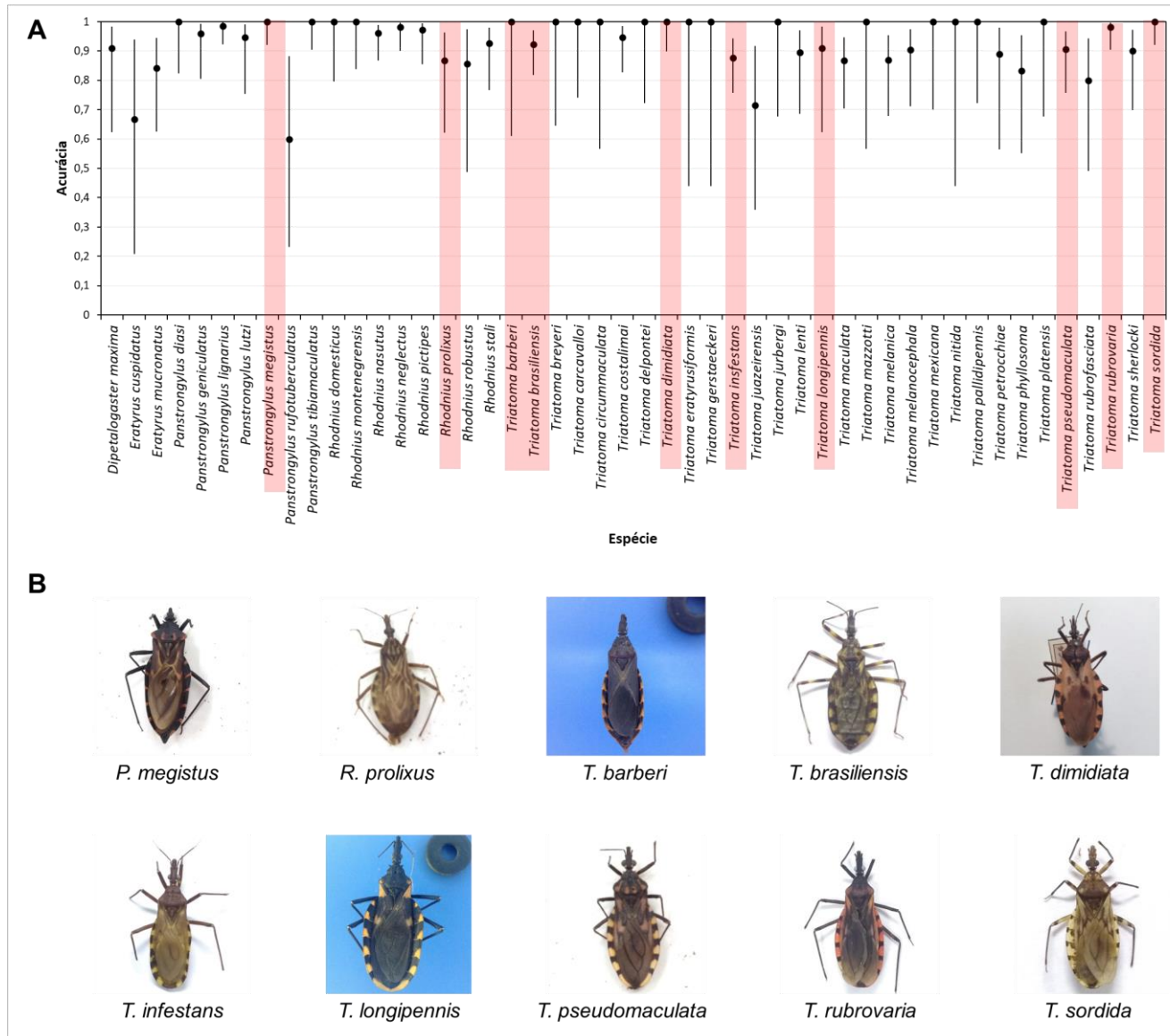


Figura 7. A – Acurácia específica da rede AlexNet para identificação de imagens dorsais de espécimes adultos pertencentes a 48 espécies de triatomíneos com intervalos de confiança de Wilson de 95%. As faixas coloridas representam as espécies representadas na figura. B – Exemplos de algumas espécies importantes na transmissão de *T. cruzi*. As imagens seguem a mesma sequência de distribuição dos pontos no gráfico.

4. Discussão

Neste estudo, mostramos que a rede neural convolucional AlexNet tem uma acurácia de ~0,93 para identificar 65 espécies de triatomíneos a partir de imagens dorsais de espécimes adultos com diversas qualidades de resolução, incluindo aquelas obtidas de câmeras de aparelhos celulares comuns (ver capítulo 1). O desempenho foi melhorado quando foram consideradas apenas as 48 espécies do banco de dados, as quais já haviam sido reportadas em domicílios e infectadas por *T. cruzi* (ver Abad-Franch & Gurgel-Gonçalves, 2023); a acurácia da rede AlexNet para identificação das espécies sinantrópicas e colonizadoras com maior capacidade vetorial (Dujardin *et al.* 2002, Waleckx *et al.* 2015, Abad-Franch 2016, Ribeiro Jr *et al.* 2021) foi > 0,94. Isso nos mostra um cenário promissor para o desenvolvimento de um sistema automatizado de identificação que possa auxiliar na vigilância dos vetores da doença de Chagas com participação comunitária em larga escala, o que poderá contribuir com a identificação das espécies, dentro da perspectiva da ciência cidadã.

O primeiro sistema de identificação automática de triatomíneos desenvolvido (Gurgel-Gonçalves *et al.*, 2017) foi fundamental para iniciarmos uma busca por ferramentas de aprendizado de máquina cada vez mais sofisticadas para desempenhar tarefas complexas de identificação de espécies, de forma automatizada e de fácil utilização, quando comparado com as chaves taxonômicas convencionais. A exploração de métodos de identificação automatizada de triatomíneos se iniciou com pesquisas básicas, a fim de entender a complexidade dos bancos de dados e dos algoritmos, quando submetidos a diversas situações e tarefas. A partir do banco de dados inicial, surgiram oportunidades para novas pesquisas com objetivo de aplicar outras técnicas de aprendizado de máquina, como a aprendizagem profunda (Khalighifar *et al.*, 2019) que melhorou as porcentagens de identificações corretas em relação aos classificadores estatísticos (Gurgel-Gonçalves *et al.*, 2017). Posteriormente, usando seus próprios bancos de imagens ou parte das nossas imagens que já haviam sido disponibilizadas, outras técnicas obtiveram bons resultados para a identificação automática desses vetores (Cruz *et al.*, 2020; Abdelghani *et al.*, 2021; Parsons & Banitaan, 2021; Cochero *et al.* 2022). Entretanto, o nosso estudo traz uma representatividade maior em número de imagens de entrada (e.g. diferentes resoluções, cores de fundo na imagem) e diversidade de espécies, contribuindo para a construção de um grande banco de imagens que ficará disponível para ser testado no futuro por outras técnicas de aprendizado de máquina. Além disso,

forneçamos informações relevantes sobre a aplicação de uma técnica de aprendizado de máquina que foi usada em uma tarefa de alta complexidade, ou seja, com um grande número de classes incluídas, quando comparada a trabalhos anteriores (Abdelghani *et al.*, 2021, Cochero *et al.* 2022, capítulo 1).

A rede AlexNet vem sendo utilizada com frequência para identificação de insetos-praga (e.g. Liu *et al.*, 2016; Cheng *et al.*, 2017; Thenmozhi & Reddy, 2019; Morankar *et al.*, 2020; Wang, 2022) e esporadicamente na identificação de macroinvertebrados aquáticos e gêneros de formigas (e.g. Raitoharju *et al.*, 2016; Marques *et al.*, 2018). No entanto, sua aplicação na identificação de insetos vetores de doenças ainda é pouco explorada. A AlexNet foi utilizada na identificação de mosquitos (Motta *et al.*, 2019), porém, obteve desempenho inferior (52,4%) em relação a outra técnica de aprendizagem profunda utilizada no mesmo estudo (GoogLeNet). Mesmo com o aumento da diversidade de espécies no nosso novo banco de dados, os resultados mostraram que o desempenho da rede AlexNet foi melhorado consideravelmente na identificação de triatomíneos, quando comparados aos estudos anteriores, com aumento de acurácia variando de 1,8% a 53,3% para 39 das 51 espécies incluídas em Gurgel-Gonçalves *et al.* (2017). No entanto, obtivemos desempenho negativo para 10 espécies, variando de -50% a -3%. Já quando comparamos com os resultados de Khalighifar *et al.* (2019), melhoramos os resultados com uma variação de 0,2% a 26,2% de acurácia para 33 espécies das 48 consideradas no estudo (Apêndice 4, Tabela 3). Contudo, houve um desempenho negativo no nosso estudo para 14 espécies, variando de -50% a -0,3%. Em comparação com os dois estudos, *P. rufotuberculatus* teve um desempenho negativo de 40% em ambos, sendo confundido com *P. lignarius* e *P. megistus* em nosso estudo. No entanto, uma explicação para este resultado é que em Gurgel-Gonçalves *et al.* (2017) essa espécie foi comparada com apenas 11 espécies mexicanas de *Triatoma*, enquanto no presente estudo comparamos com 64 espécies de 7 gêneros. No entanto, a nossa análise contou com uma adição de 320 e 156 imagens para essas espécies, respectivamente, o que pode ter contribuído para diminuição do desempenho da rede, devido ao desbalanceamento das classes, favorecendo as confusões das identificações. Já *T. carcavallo* (com média de desempenho negativo de ~30% em relação aos dois estudos) foi mais frequentemente confundido com *T. rubrovaria* e *T. pintodiasi*, confusão que pode ter sido influenciada pela sua similaridade morfológica, já que segundo a revisão filogenética que consideramos no

nosso estudo, essas três espécies pertencem ao complexo *T. rubrovaria* (Monteiro *et al.*, 2018). O mesmo acontece com *T. platensis* que foi confundido com *T. delpontei* (com média de ~10%), e que estão incluídas no grupo de espécies *T. infestans*; *T. juazeirensis* com *T. melanica* (média de ~24%), no complexo *T. brasiliensis*; *T. guazu* com *T. rubrovaria*, *T. vandae* e *T. williami* (média de ~21%) no clado *T. sordida*; e, as espécies mexicanas, como *T. gerstaeckeri* sendo confundida como *T. mexicana* e *T. phyllosoma* (-50%, em comparação aos dois estudos), pertencentes ao grupo *T. phyllosoma* (Monteiro *et al.*, 2018). No entanto, neste último caso, o conjunto de imagens de *T. gerstaeckeri* (14 imagens) foi muito inferior aos de *T. mexicana* e *T. phyllosoma*, 46 e 60 imagens, respectivamente, o que pode estar relacionado com a mesma situação do *P. rufotuberculatus* explicada anteriormente.

Observando os resultados exploratórios, e considerando que há a possibilidade de o tamanho da amostra para cada classe (número de fotos) melhorar a acurácia específica para cada espécie, sugerimos que essa não seja a única característica que influencia a discriminação das classes. As características morfológicas de cada espécie, sejam do mesmo gênero ou gêneros diferentes, podem ter efeitos mais significativos na identificação das espécies do que a quantidade de imagens para cada classe. Algumas espécies provavelmente têm menor acurácia específica devido à baixa quantidade de imagens incluídas no treinamento da rede, por exemplo, *C. pilosa*, *P. rufotuberculatus*, *T. eratyrisiformis*, *T. gerstaeckeri*, *T. guazu* [neste último, apesar de recentemente ter sido considerado como sinônimo de *T. williami* (Oliveira Correia *et al.*, 2022), das três imagens confundidas apenas uma foi identificada como *T. williami*]. Entretanto, outras espécies são tão morfológicamente distintas, que esse baixo número de imagens parece não interferir na predição da rede. Por outro lado, também houve espécies com um número alto de imagens incluídas, mas com desempenho menor na acurácia específica (por exemplo, *Ps. coreodes*, que foi frequentemente confundido com *Ps. tertius*), que pode ter acontecido devido à grande semelhança morfológica entre essas duas espécies (Oliveira *et al.*, 2023). No futuro pretendemos implementar uma função que mostre qual região das imagens de entrada é mais usada pelo algoritmo para realizar as predições das classes (e.g. Park *et al.*, 2020; Couret *et al.*, 2023).

Em nossa análise com 65 espécies, 43 delas apresentaram acurácia específica igual ou superior a 90% e destas, 21 espécies foram identificadas com 100% de acurácia. Na análise que considerou apenas as 48 espécies encontradas em

casas e infectadas por *T. cruzi*, 36 delas apresentaram acurácia específica superior a 90%, e 20 foram identificadas com 100% de acurácia. Apesar do desempenho da rede AlexNet ter sido um pouco melhor para 48 espécies em relação as 65 espécies, essa diferença foi apenas de 1,8%. Das espécies que atingiram 90% ou mais de acurácia específica, 29 delas foram comuns nas duas análises, enquanto 9 espécies comuns atingiram 100%. Como discutido anteriormente, apesar de não termos realizado uma análise mais rigorosa sobre o melhor e o pior desempenho da rede AlexNet para algumas espécies, inferimos que o número de imagens de treinamento de cada classe não foi o único determinante para se obter boas ou piores predições, uma vez que espécies com baixo número de imagens de treinamento (por exemplo, *E. cuspidatus*, *T. barberi*, *T. jurbergi*, *T. mexicana*, *T. nitida*) obtiveram 100% de acurácia específica, enquanto outras com um número relativamente maior de imagens de treinamento (por exemplo, *Ps. coreodes*, *T. carcavalloii*, *T. phyllosoma*, *T. matogrossensis*) obtiveram acurácias menores até que 80% (por exemplo, *Ps. coreodes*). Essa diferença na acurácia da rede para identificar imagens de algumas espécies pode estar sofrendo um efeito maior devido às características morfológicas únicas de cada espécie, o que pode facilitar ou confundir as identificações previstas pela rede (ver características morfológicas em Lent & Wygodzinsky, 1979).

Na análise com 48 espécies, observamos que as acurácias para espécies mais relevantes para saúde pública apresentaram altos valores (*P. megistus*, *R. prolixus*, *T. barberi*, *T. brasiliensis*, *T. dimidiata*, *T. infestans*, *T. longipennis*, *T. pseudomaculata*, *T. rubrovaria* e *T. sordida*), variando de 0,867 a 1,000. Neste caso, *R. prolixus* e *T. infestans* apresentaram os menores valores de acuraria específica, *P. megistus*, *T. barberi*, *T. dimidiata* e *T. sordida* alcançaram valores máximos e os demais tiveram valores superiores a 90%. Esses resultados mostram mais um avanço na identificação automática de triatomíneos em nível de espécie. Atualmente os sistemas de identificação baseados em imagens como Google Lens e Bug ID não permitem uma identificação específica acurada, algumas vezes confundindo triatomíneos com besouros e dificilmente identificando as espécies corretamente (RGG; VLM, observação pessoal), o que deve estar relacionado a uma base de dados incompleta, fragmentada e não validada por especialistas. Os sistemas de identificação automática devem ser acurados principalmente na identificação dessas espécies mais relevantes para que a vigilância funcione adequadamente. Por exemplo, na América do Sul, *P. megistus* é uma espécie de triatomíneo que pode

ocupar ecótopos artificiais, como galinheiros próximo às residências, e consequentemente se tornar domiciliada, aumentando o risco da transmissão de *T. cruzi* para humanos (Forattini *et al.*, 1977a; 1977b; 1978). Já *P. tibiamaculatus* dificilmente coloniza as casas (Ribeiro Jr *et al.*, 2015). Nesta situação, as ações da vigilância e controle devem ser diferentes para espécies domiciliadas e não domiciliadas (Abad-Franch 2016, Dias *et al.*, 2016). Se o sistema de identificação automática confunde essas espécies, a vigilância e controle podem não ser efetivos. O mesmo ocorre com as espécies do complexo *T. brasiliensis* (Costa *et al.*, 2003). *T. brasiliensis* é um dos vetores mais importantes nas zonas semiáridas do Nordeste do Brasil devido às suas características biológicas e ecológicas, como sua alta densidade populacional e sua capacidade de invadir e colonizar residências (Costa, 1999; Costa *et al.*, 2003; Sarquis *et al.*, 2006); no entanto, essas características não são compartilhadas por outros integrantes do complexo, como *T. petrocchiaae*, por exemplo. Outra espécie também considerada muito importante é *T. sordida*, que está bem adaptada a ecótopos artificiais, sendo frequentemente registrada invadindo residências e infectada por *T. cruzi* e com uma ampla distribuição no Brasil (Gurgel-Gonçalves *et al.*, 2012) e outros países como Argentina, Bolívia, Paraguai e Uruguai (Forattini *et al.*, 1971; Schweigmann *et al.*, 1997; Bar *et al.*, 2001; Madeira *et al.*, 2021; Carcavallo *et al.* 1997). E para a América do Norte e Central (apesar da sua distribuição ultrapassar esses limites) destacamos *T. dimidiata*, que é o principal vetor da doença de Chagas na Guatemala, El Salvador, Nicarágua e Costa Rica e pode apresentar hábitat que inclui ecótopos silvestres, peridomésticos e domésticos (Petana, 1971; Zeledón *et al.*, 1973; Monroy *et al.*, 2003; Dorn *et al.*, 2007). Dessa forma, nosso estudo apresenta boas perspectivas para a identificação dos triatomíneos, principalmente para as espécies mais importantes do ponto de vista da saúde pública.

Apesar deste estudo ser um grande avanço na identificação dos vetores da doença de Chagas, nossa base de dados representa ~50% das espécies atualmente descritas de triatomíneos, principalmente espécies que ocorrem no Brasil e México. Assim, há necessidade de ampliar o número de observações e espécies analisadas para tornar o banco de dados mais representativo, incluindo espécies de outros países das Américas. No entanto, se considerarmos apenas as espécies mais encontradas em residências no Brasil (11 espécies, Silveira & Dias, 2011, Dias *et al.*, 2016) e no México (19 espécies, Ramsey *et al.* 2015), o banco de dados de imagens de

triatomíneos inclui 100% e 80% das espécies no Brasil e no México, respectivamente. Dessa forma, considerando a lista de espécies com maior relevância em saúde pública, com registro em casas e infectadas com *T. cruzi* (Abad-Franch e Gurgel-Gonçalves, 2021), existem 56 espécies sinantrópicas, e no nosso estudo foram incluídas 48 espécies (~86%). A execução do estudo também evidenciou a necessidade de melhorar as condições das coleções de triatomíneos, pois a confirmação taxonômica de alguns espécimes foi dificultada porque eles não estavam rotulados ou estavam em um estado mal preservado. Futuras colaborações de pesquisa com Peru, Equador e Panamá serão fundamentais para obtenção de imagens de *R. ecuadoriensis* (relevante no Equador e Peru), *P. chinai* (no Peru), *R. pallescens* (Panamá), por exemplo.

Após a incorporação de um número maior de espécies, o sistema deverá ser disponibilizado por meio de um aplicativo de acesso aberto. A ideia é organizar um sistema de identificação automática de triatomíneos das Américas em que qualquer pessoa com um smartphone conectado à internet possa fotografar um triatomíneo, enviar a foto para o sistema e receber a identificação do inseto junto com informações sobre sua importância como vetor. Outro passo importante nessa direção foi o desenvolvimento do GeoVin (Cochero *et al.*, 2022), que coleta informações geográficas sobre triatomíneos argentinos por meio de relatos de cidadãos sobre descobertas de triatomíneos. Os dados de ocorrência (fotos e coordenadas geográficas) são enviados online, armazenados e automaticamente integrados ao conjunto de dados de ocorrência do GeoVin. Aplicações baseadas em sistemas automatizados de identificação podem fortalecer substancialmente a vigilância de vetores da doença de Chagas, ao mesmo tempo em que promovem a ciência cidadã. Acreditamos que nosso trabalho contribui com novas soluções para melhorar a identificação automatizada de triatomíneos. O sistema automatizado que estamos aplicando, e conseqüentemente, desenvolvendo (por ser treinado com o nosso próprio banco de imagens) resolverá um grande problema na vigilância de vetores da doença de Chagas, que está se tornando cada vez mais fraca na América Latina, em parte porque a experiência na taxonomia de vetores de doenças está diminuindo (Casas *et al.*, 2016; Gurgel-Gonçalves, 2022). Neste cenário de sistemas de vigilância cada vez mais fracos, precisamos de um sistema de vigilância forte com envolvimento da comunidade (por exemplo, Abad-Franch *et al.*, 2011) e sistemas de identificação

automatizados para melhorar a detecção de vetores pelas famílias. Acreditamos que este será o futuro da vigilância do vetor da doença de Chagas.

5. Conclusão

Neste estudo, evidenciamos que a rede neural convolucional AlexNet, quando alimentada com um banco de imagens de treinamento amplo, diversificado e bem estruturado, exibe um desempenho altamente promissor para identificação de uma ampla variedade de espécies dos vetores da doença de Chagas. Enquanto enfrentamos a complexidade inerente à tarefa de identificação desses insetos, a AlexNet se destacou como uma ferramenta com potencial para fortalecer a vigilância entomológica, garantindo a identificação precisa e confiável dos triatomíneos. Nossos resultados ressaltam a viabilidade dessa abordagem automatizada, especialmente em termos de reconhecimento de espécies de maior relevância para a saúde pública. Espécies como *P. megistus* e *T. brasiliensis*, fundamentais na transmissão da doença de Chagas, apresentaram altos níveis de acurácia, ressaltando a capacidade da AlexNet em discernir entre esses vetores cruciais. Isso promete melhorar substancialmente a eficácia das ações de vigilância e controle, essenciais para prevenir a disseminação da doença. Embora este estudo represente um passo significativo, reconhecemos a necessidade contínua de expansão da base de dados e a inclusão de espécies de várias regiões das Américas. À medida que trabalhamos para enriquecer o banco de imagens, visamos disponibilizar um aplicativo de acesso aberto, permitindo que qualquer pessoa com um smartphone contribua para a vigilância e a identificação dos triatomíneos. Em última análise, a aplicação da AlexNet e a exploração de sistemas automatizados de identificação representam uma resposta eficaz aos desafios crescentes enfrentados pela vigilância de vetores da doença de Chagas. Acreditamos que essa abordagem, combinada com o engajamento comunitário, conduzirá a uma vigilância mais robusta e, em última instância, à contenção da propagação da doença, tornando-se o caminho promissor para o futuro da saúde pública.

6. Referências

- Abad-Franch, F. (2016). A simple, biologically sound, and potentially useful working classification of Chagas disease vectors. *Memórias do Instituto Oswaldo Cruz*, 111, 649-651.
- Abad-Franch, F., & Gurgel-Gonçalves, R. (2021). The ecology and natural history of wild Triatominae in the Americas. In Guarneri, A & Lorenzo, M (Ed.). *Triatominae - The Biology of Chagas Disease Vectors*. (Vol. 5, Chap. 16, pp. 387-445). Cham: Springer International Publishing.
- Abdelghani, B. A., Banitaan, S., Maleki, M., & Mazen, A. (2021). Kissing bugs identification using convolutional neural network. *IEEE Access*, 9, 140539-140548.
- Bar, M. E., & Wisnivesky-Colli, C. (2001). *Triatoma sordida* Stål 1859 (Hemiptera, Reduviidae: Triatominae) in palms of northeastern Argentina. *Memórias do Instituto Oswaldo Cruz*, 96, 895-899.
- Carvalho, M. R., Bockmann, F. A., Amorim, D. S., Brandão, C. R. F., Vivo, M., Figueiredo, J. L., Britski, H. A., Pinna, M. C. C., Menezes, N. A., Marques, F. P. L., Papavero, N., Cancellato, E. M., Crisci, J. V., McEachran, J. D., Schelly, R., Lundberg, J. G., Gill, A. C., Britz, R., Wheeler, Q. D., Stiasny, M. L. J., Parenti, L. R., Page, L. M., Wheeler, W. C., Faivovich, J., & Nelson, G. J. (2007). Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, 34, 140-143.
- Carcavallo, R. U., Galindez, I., José, J., & Herman, L. (1997). Atlas dos vetores da doença de Chagas nas Américas. Rio de Janeiro: Fiocruz Editorial.
- Casas, J., Lazzari, C., Insausti, T., Launois, P., & Fouque, F. (2016). Mapping of courses on vector biology and vector-borne diseases systems: time for a worldwide effort. *Memórias do Instituto Oswaldo Cruz*, 111, 717-719.
- Cochero, J., Patteri, L., Balsalobre, A., Ceccarelli, S., & Marti, G. (2022). A convolutional neural network to recognize Chagas disease vectors using mobile phone images. *Ecological Informatics*, 68, 101587.
- Costa, J., Almeida, C. E., Dotson, E. M., Lins, A., Vinhaes, M., Silveira, A. C., & Beard, C. B. (2003). The epidemiologic importance of *Triatoma brasiliensis* as a Chagas

- disease vector in Brazil: a revision of domiciliary captures during 1993-1999. *Memórias do Instituto Oswaldo Cruz*, 98, 443-449.
- Costa, J. (1999). The synanthropic process of Chagas disease vectors in Brazil, with special attention to *Triatoma brasiliensis* Neiva, 1911 (Hemiptera, Reduviidae, Triatominae) population, genetical, ecological, and epidemiological aspects. *Memórias do Instituto Oswaldo Cruz*, 94, 239-241.
- Couret, J., Moreira, D. C., Bernier, D., Loberti, A. M., Dotson, E. M., & Alvarez, M. (2020). Delimiting cryptic morphological variation among human malaria vector species using convolutional neural networks. *PLoS neglected tropical diseases*, 14(12), e0008904.
- Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, 141, 351-356.
- Cruz, D. D., Arellano, E., Denis Ávila, D., & Ibarra-Cerdeña, C. N. (2020). Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: a case for the cryptic dimidiata complex. *Parasites & vectors*, 13, 1-12.
- Dias, J. C. P., Ramos Jr, A. N., Gontijo, E. D., Luquetti, A., Shikanai-Yasuda, M. A., Coura, J. R., ... & Alves, R. V. (2016). II Consenso Brasileiro em doença de Chagas, 2015. *Epidemiologia e Serviços de Saúde*, 25, 7-86.
- Dorn, P. L., Monroy, C., & Curtis, A. (2007). *Triatoma dimidiata* (Latreille, 1811): A review of its diversity across its geographic range and the relationship among populations. *Infection, Genetics and Evolution*, 7(2), 343-352.
- Dujardin, J. P., Schofield, C. J., Panzera, F., & Matias, A. (2002). *Los vectores de la enfermedad de Chagas* (pp. 25-64). Bruxelles: Académie Royale des Sciences d'Outre-Mer.
- Forattini, O. P., Ferreira, O. A., Silva, E. O. D. R. E., & Rabello, E. X. (1977). Aspectos ecológicos da tripanossomíase americana: VIII-Domiciliação de *Panstrongylus megistus* e sua presença extradomiciliar. *Revista de Saúde Pública*, 11(1), 73-86.
- Forattini, O. P., Silva, E. O. D. R., Ferreira, O. A., Rabello, E. X., Santos, J. L. F., & Lima, A. R. D. (1977). Aspectos ecológicos da tripanossomíase americana: XI-

- Domiciliação de *Panstrongylus megistus* e potencial enzoótico. *Revista de Saúde Pública*, 11, 527-550.
- Forattini, O. P., Rocha e Silva, E. O. D., Rabello, E. X., Andrade, J. C. R. D., & Rodrigues, V. L. C. C. (1978). Ecological aspects of South American Trypanosomiasis: XIII-Domestic enzootic potencial in *Panstrongylus megistus* area submitted to epidemiological surveillance. *Revista de Saúde Pública*, 12, 417-424.
- Forattini, O. P., Rocha e Silva, E. O. D., Ferreira, O. A., Rabello, E. X., & Pattoli, D. G. (1971). Aspectos ecológicos da tripanossomose americana: III-Dispersão local de triatomíneos, com especial referência ao *Triatoma sordida*. *Revista de Saúde Pública*, 5, 193-205.
- Galvão, C. (2014). Vetores da doença de Chagas no Brasil. Curitiba: Sociedade Brasileira de Zoologia. Curitiba: Sociedade Brasileira de Zoologia.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT press.
- Gurgel-Gonçalves, R., Galvao, C., Costa, J., & Peterson, A. T. (2012). Geographic distribution of Chagas disease vectors in Brazil based on ecological niche modeling. *Journal of tropical medicine*, 2012, 705326.
- Gurgel-Gonçalves, R., Komp, E., Campbell, L. P., Khalighifar, A., Mellenbruch, J., Mendonça, V. J., Owens, H. L., Felix, K. L. C., Peterson, A. T., & Ramsey, J. M. (2017). Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the Virtual Vector Lab. *PeerJ*, 5, e3040.
- Gurgel-Gonçalves, R., Abad-Franch, F., de Almeida, M. R., Obara, M. T., de Souza, R. D. C. M., Batista, J. A. D. S., & Rocha, D. D. A. (2021). TriatoDex, an electronic identification key to the Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease: Development, description, and performance. *Plos one*, 16(4), e0248628.
- Gurgel-Gonçalves, R. (2023). Stronger control-surveillance systems for vector-borne Chagas disease. *Memórias do Instituto Oswaldo Cruz*, 117, e210130chgsb..
- Harrell, F., & Dupont, C. (2023). Package 'Hmisc': Harrell miscellaneous. Disponível em <http://www.who.int/mediacentre/factsheets/fs340/en/>. Acesso em: 19 jul. 2023.

- Høye, T. T., Ärje, J., Bjerger, K., Hansen, O. L., Iosifidis, A., Leese, F., Mann, H. M. R., Meissner, K., Melvad, C., & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, 118(2), e2002545117.
- Kittichai, V., Kaewthamasorn, M., Samung, Y., Jomtarak, R., Naing, K. M., Tongloy, T., Chuwongin, S., & Boonsang, S. (2023). Automatic identification of medically important mosquitoes using embedded learning approach-based image-retrieval system. *Scientific Reports*, 13(1), 10609.
- Khalighifar, A., Komp, E., Ramsey, J. M., Gurgel-Gonçalves, R., & Peterson, A. T. (2019). Deep learning algorithms improve automated identification of Chagas disease vectors. *Journal of medical entomology*, 56(5), 1404-1410.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lent, H., & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bulletin of the American museum of Natural History*, 163(3), 123-520.
- Likith, S., Reddy, B. R., & Reddy, K. S. (2021). A Smart System for Detection and Classification of Pests Using YOLO AND CNN Techniques. In *2021 International Conference on Computational Performance Evaluation (ComPE)* (pp. 049-052). IEEE.
- Liu, Z., Gao, J., Yang, G., Zhang, H., & He, Y. (2016). Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific reports*, 6(1), 20410.
- Marques, A. C. R., M. Raimundo, M., B. Cavalheiro, E. M., FP Salles, L., Lyra, C., & J. Von Zuben, F. (2018). Ant genera identification using an ensemble of convolutional neural networks. *Plos one*, 13(1), e0192011.
- Madeira, F. F., Delgado, L. M. G., Bittinelli, I. D. F., de Oliveira, J., Ravazi, A., Dos Reis, Y. V., Oliveira, A. B. B., Cristal, D. C., Galvão, C., Azeredo-Oliveira, M. T. V., Rosa, J. A., & Alevi, K. C. C. (2021). *Triatoma sordida* (Hemiptera, Triatominae) from La

- Paz, Bolivia: an incipient species or an intraspecific chromosomal polymorphism?. *Parasites & Vectors*, 14, 1-8.
- Monteiro, F. A., Weirauch, C., Felix, M., Lazoski, C., & Abad-Franch, F. (2018). Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. *Advances in parasitology*, 99, 265-344.
- Monroy, C., Rodas, A., Mejia, M., & Tabaru, Y. (1998). Wall plastering and paints as methods to control vectors of Chagas disease in Guatemala. *Medical Entomology and Zoology*, 49(3), 187-193.
- Morankar, D., Shinde, D. M., & Pawar, S. R. (2020). Identification of pests and diseases using alex-net. *SSRN Electronic Journal*, 7(4), 53-61.
- Motta, D., Santos, A. Á. B., Winkler, I., Machado, B. A. S., Pereira, D. A. D. I., Cavalcanti, A. M., Fonseca, E. O. L., Kirchner, F., & Badaró, R. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. *PloS one*, 14(1), e0210829.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17(8), 857-872.
- Oliveira, J., Alevi, K. C. C., Almeida, C. E., Olaia, N., Cacini, G. L., Galvão, C., Herrera, H. M., Santos, F. M., & Rosa, J. A. D. (2023). Exploring the Hidden World of Vectors of Chagas Disease: A Fascinating Look at the Taxonomic Aspects of the *Psammolestes* Genus (Hemiptera, Triatominae). *Life*, 13(5), 1081.
- Oliveira Correia, J. P. S., Gil-Santana, H. R., Dale, C., & Galvão, C. (2022). *Triatoma guazu* Lent and Wygodzinsky is a junior synonym of *Triatoma williami* Galvão, Souza and Lima. *Insects*, 13(7), 591.
- Ong, S. Q., Nair, G., Yusof, U. K., & Ahmad, H. (2022). Community-based mosquito surveillance: an automatic mosquito-on-human-skin recognition system with a deep learning algorithm. *Pest Management Science*, 78(10), 4092-4104.
- Park, J., Kim, D. I., Choi, B., Kang, W., & Kwon, H. W. (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Scientific reports*, 10(1), 1012.

- Parsons, Z., & Banitaan, S. (2021). Automatic identification of Chagas disease vectors using data mining and deep learning techniques. *Ecological Informatics*, 62, 101270.
- Petana, W. B. (1971). American trypanosomiasis in British Honduras: X: Natural habitats and ecology of *Triatoma dimidiata* (Hemiptera, Reduviidae) in the El Cayo and Toledo Districts, and the prevalence of infection with *Trypanosoma (Schizotrypanum) cruzi* in the wild-caught bugs. *Annals of Tropical Medicine & Parasitology*, 65(2), 169-178.
- Posit Software. (2022). RStudio 2023.03.1.446. <https://posit.co/products/open-source/rstudio/>.
- Raitoharju, J., Riabchenko, E., Meissner, K., Ahmad, I., Iosifidis, A., Gabbouj, M., & Kiranyaz, S. (2016). Data enrichment in fine-grained classification of aquatic macroinvertebrates. In *2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)* (pp. 43-48). IEEE.
- R Core Team. (2022). R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ribeiro Jr, G., Gurgel-Goncalves, R., Reis, R. B., Santos, C. G. S. D., Amorim, A., Andrade, S. G., & Reis, M. G. (2015). Frequent house invasion of *Trypanosoma cruzi*-infected triatomines in a suburban area of Brazil. *PLoS Neglected Tropical Diseases*, 9(4), e0003678.
- Ribeiro-Jr, G., Abad-Franch, F., Sousa, O. M., Santos, C. G., Fonseca, E. O. L., Santos, R. F., Cunha, G. M., Carvalho, C. M. M., Reis, R. B., Gurgel-Gonçalves, R., & Reis, M. G. (2021). TriatoScore: an entomological-risk score for Chagas disease vector control-surveillance. *Parasites & Vectors*, 14, 1-15.
- Sarquis, O., Sposina, R., Oliveira, T. G. D., Mac Cord, J. R., Cabello, P. H., Borges-Pereira, J., & Lima, M. M. (2006). Aspects of peridomiliary ecotopes in rural areas of Northeastern Brazil associated to triatomine (Hemiptera, Reduviidae) infestation, vectors of Chagas disease. *Memórias do Instituto Oswaldo Cruz*, 101, 143-147.
- Schweigmann, N., Pietrokovsky, S., Conti, O., Escosteguy, M., Bottazzi, V., Solarz, N., & Wisnivesky-Colli, C. (1997). Infection of *Triatoma guasayana*, *Triatoma sordida*

- and *Triatoma infestans* by *Trypanosoma cruzi* from a naturally infected opossum. *Memórias do Instituto Oswaldo Cruz*, 92, 151-152.
- Silveira, A. C., & Dias, J. C. P. (2011). O controle da transmissão vetorial. *Revista da Sociedade Brasileira de Medicina Tropical*, 44, 52-63.
- Thenmozhi, K., & Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 164, 104906.
- Valan, M., Makonyi, K., Maki, A., Vondráček, D., & Ronquist, F. (2019). Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology*, 68(6), 876-895.
- Waleckx, E., Gourbière, S., & Dumonteil, E. (2015). Intrusive versus domiciliated triatomines and the challenge of adapting vector control practices against Chagas disease. *Memórias do Instituto Oswaldo Cruz*, 110, 324-338.
- Wang, B. (2022). Identification of crop diseases and insect pests based on deep learning. *Scientific Programming*, 2022, 1-10.
- WHO – World Health Organization. (2021). *Chagas disease (American trypanosomiasis)*. Disponível em https://www.who.int/health-topics/chagasdisease#tab=tab_1.
- Zeledón, R., Solano, G., Zúñiga, A., & Swartzwelder, J. C. (1973). Biology and ethology of *Triatoma dimidiata* (Latreille, 1811). III. Habitat and blood sources. *Journal of medical entomology*, 10(4), 363-370.
- Zhao, Y., Fan, M., Li, H., & Cai, W. (2023). Review of Kissing Bugs (Hemiptera: Reduviidae: Triatominae) from China with Descriptions of Two New Species. *Insects*, 14(5), 450.

APÊNDICE 1

Apêndice 1. Número total de imagens dorsais de espécimes adultos de triatomíneos usadas no treinamento e teste da rede AlexNet considerando o presente estudo (Miranda, 2023) e o banco de dados disponível (Gurgel-Gonçalves *et al.*, 2017).

Espécies	Nº de imagens		
	Miranda, 2023	Gurgel-Gonçalves <i>et al.</i> , 2017	Total
<i>Cavernicola lenti</i>	27	32	59
<i>Cavernicola pilosa</i>	17		17
<i>Dipetalogaster maxima</i>	54		54
<i>Eratyrus cuspidatus</i>	14		14
<i>Eratyrus mucronatus</i>	81	11	92
<i>Panstrongylus diasi</i>	60	30	90
<i>Panstrongylus geniculatus</i>	79	45	124
<i>Panstrongylus lignarius</i>	320	30	350
<i>Panstrongylus lutzi</i>	62	35	96
<i>Panstrongylus megistus</i>	156	71	228
<i>Panstrongylus rufotuberculatus</i>	12	13	25
<i>Panstrongylus tibiamaculatus</i>	139	41	180
<i>Psammolestes coreodes</i>	137		137
<i>Psammolestes tertius</i>	134	36	170
<i>Rhodnius brethesi</i>	128	28	156
<i>Rhodnius domesticus</i>	43	30	73
<i>Rhodnius marabaensis</i>	52		52
<i>Rhodnius milesi</i>	14	43	57
<i>Rhodnius montenegrensis</i>	56	43	99
<i>Rhodnius nasutus</i>	173	80	253
<i>Rhodnius neglectus</i>	198	67	265
<i>Rhodnius pictipes</i>	119	58	177
<i>Rhodnius prolixus</i>	72		72
<i>Rhodnius robustus</i>	34		34
<i>Rhodnius stali</i>	137		137
<i>Triatoma arthurneivai</i>	13	32	45
<i>Triatoma baratai</i>	35	29	64
<i>Triatoma barberi</i>		31	31
<i>Triatoma brasiliensis</i>	193	66	259
<i>Triatoma breyeri</i>	35		35
<i>Triatoma carcavalloii</i>	15	38	53
<i>Triatoma circummaculata</i>		22	22
<i>Triatoma costalimai</i>	125	64	189
<i>Triatoma delpontei</i>	19	31	50
<i>Triatoma dimidiata</i>	55	114	169
<i>Triatoma eratyrusiformis</i>	14		14
<i>Triatoma gerstaeckeri</i>	2	12	14
<i>Triatoma guazu</i>	3	29	32
<i>Triatoma infestans</i>	191	55	246
<i>Triatoma juazeirensis</i>	13	21	34
<i>Triatoma jurbergi</i>	41		41
<i>Triatoma klugi</i>	40		40
<i>Triatoma lenti</i>	57	40	97
<i>Triatoma longipennis</i>		52	52
<i>Triatoma maculata</i>	110	40	150

<i>Triatoma matogrossensis</i>	69	33	102
<i>Triatoma mazzottii</i>		23	23
<i>Triatoma melanica</i>	84	29	113
<i>Triatoma melanocephala</i>	73	31	104
<i>Triatoma mexicana</i>		46	46
<i>Triatoma nitida</i>		16	16
<i>Triatoma pallidipennis</i>	6	43	49
<i>Triatoma petrocchiai</i>	42		42
<i>Triatoma phyllosoma</i>		60	60
<i>Triatoma pintodiasi</i>	50	26	76
<i>Triatoma platensis</i>	13	28	41
<i>Triatoma pseudomaculata</i>	105	56	161
<i>Triatoma rubrofaciata</i>	9	42	51
<i>Triatoma rubrovaria</i>	217	55	272
<i>Triatoma sherlocki</i>	69	32	101
<i>Triatoma sordida</i>	120	104	224
<i>Triatoma vanda</i>	45	30	75
<i>Triatoma vitticeps</i>	79	48	127
<i>Triatoma williami</i>	16	20	36
<i>Triatoma wygodzinsky</i>	30		30
Total	4306	2091	6397

APÊNDICE 2

Apêndice 2. Protocolo utilizado para a configuração da rede AlexNet para identificação de triatomíneos

1. Preparação dos Dados

- Foi criada uma pasta chamada "Saida" na mesma localização do algoritmo para armazenar as imagens utilizadas neste estudo.
- Utilizou-se a função `imageDatastore` para criar um objeto `imageDatastore` denominado 'imds'. Esse objeto foi construído com base nas imagens contidas na pasta 'Saida', incluindo as imagens presentes em subpastas, e os nomes das pastas foram utilizados como rótulos para cada imagem.

2. Divisão do Conjunto de Dados

- Os rótulos únicos das imagens foram calculados para determinar o número de classes presentes no conjunto de dados.
- Para a divisão adequada do conjunto de dados, utilizou-se a função `splitEachLabel`. Essa função dividiu o objeto `imageDatastore` 'imds' em três conjuntos distintos: Treinamento, Validação e Teste.
- O conjunto de Treinamento foi composto por 70% das imagens disponíveis, o conjunto de Teste contém 20% das imagens e o conjunto de Validação abrange os 10% restantes. A divisão foi realizada de forma aleatória para garantir a representatividade dos dados em cada conjunto.

3. Carregamento da Rede Pré-Treinada AlexNet

- A arquitetura da rede neural convolucional pré-treinada AlexNet foi carregada utilizando a função 'alexnet'. As camadas da rede foram armazenadas na variável 'layers'.
- Uma nova camada totalmente conectada (`fullyConnectedLayer`) foi adicionada à rede, projetada para o número específico de classes identificadas anteriormente.
- Para adaptar a rede ao problema de classificação multiclasse, a camada de classificação original (camada 23) da AlexNet foi substituída pela nova camada totalmente conectada.

- Por fim, a última camada de classificação também foi ajustada para corresponder ao número total de classes presentes no conjunto de dados.

4. Definição das Opções de Treinamento

- Utilizou-se a função `trainingOptions` para estabelecer as opções de treinamento da rede neural.
- O otimizador escolhido para o treinamento foi o 'stochastic gradient descent with momentum' (sgdm).
- Parâmetros importantes foram configurados, incluindo a taxa de aprendizado inicial (0.001) e o tamanho do lote de treinamento (128).
- A GPU foi habilitada para acelerar o processo de treinamento, caso esteja disponível.
- Gráficos de progresso do treinamento, como 'plots' e 'training-progress', foram gerados para auxiliar a análise do desempenho da rede durante o processo de treinamento.
- A quantidade máxima de épocas de treinamento foi definida em 100.
- O conjunto de Validação (conjunto de validação interno) foi utilizado para monitorar o desempenho da rede durante o treinamento.

APÊNDICE 3

Apêndice 3. Erros nas identificações de imagens dorsais de espécimes adultos de espécies de triatomíneos preditas pela rede AlexNet.

ID foto	Espécie	Predição
Cavernicola_pilosa_IRR3.jpg	<i>Cavernicola pilosa</i>	<i>Cavernicola lenti</i>
Eratyrus_mucronatus61.jpg	<i>Eratyrus mucronatus</i>	<i>Triatoma infestans</i>
Panstrongylus_lignarius_IRR26.jpg	<i>Panstrongylus lignarius</i>	<i>Panstrongylus geniculatus</i>
Panstrongylus_rufotuberculatus_IRR2.jpg	<i>Panstrongylus rufotuberculatus</i>	<i>Panstrongylus lignarius</i>
Panstrongylus_rufotuberculatus_IRR4.jpg	<i>Panstrongylus rufotuberculatus</i>	<i>Panstrongylus megistus</i>
Psammolestes_coreodes_IRR4.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_coreoides_SP1_12.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_coreoides_SP1_129.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_coreoides_SP1_38.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_coreoides_SP1_62.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_coreoides_SP1_86.jpg	<i>Psammolestes coreodes</i>	<i>Psammolestes tertius</i>
Psammolestes_tertius_SP1_104.jpg	<i>Psammolestes tertius</i>	<i>Psammolestes coreodes</i>
Psammolestes_tertius_SP1_51.jpg	<i>Psammolestes tertius</i>	<i>Psammolestes coreodes</i>
Psammolestes_tertius_SP1_66.jpg	<i>Psammolestes tertius</i>	<i>Psammolestes coreodes</i>
Rhodnius_domesticus_IRR5.jpg	<i>Rhodnius domesticus</i>	<i>Rhodnius stali</i>
Rhodnius_marabaensis_SP1_4.jpg	<i>Rhodnius marabaensis</i>	<i>Panstrongylus tibiamaculatus</i>
Rhodnius_montenegrensis_SP5_2.jpg	<i>Rhodnius montenegrensis</i>	<i>Rhodnius milesi</i>
Rhodnius_nasutus764.jpeg	<i>Rhodnius nasutus</i>	<i>Cavernicola pilosa</i>
Rhodnius_nasutus788.jpeg	<i>Rhodnius nasutus</i>	<i>Triatoma infestans</i>
Rhodnius_nasutus_IRD3.jpg	<i>Rhodnius nasutus</i>	<i>Rhodnius robustus</i>
Rhodnius_pictipes_IRD18.jpg	<i>Rhodnius pictipes</i>	<i>Rhodnius robustus</i>
Rhodnius_prolixus_IRD4.jpg	<i>Rhodnius prolixus</i>	<i>Rhodnius pictipes</i>
Rhodnius_robustus_IRD34.jpg	<i>Rhodnius robustus</i>	<i>Rhodnius prolixus</i>
Rhodnius_stali_IRR9.jpg	<i>Rhodnius stali</i>	<i>Rhodnius pictipes</i>
Triatoma_arthurneivai_IRR10.jpg	<i>Triatoma arthurneivai</i>	<i>Triatoma baratai</i>
Triatoma_arthurneivai_IRR8.jpg	<i>Triatoma arthurneivai</i>	<i>Triatoma rubrofasciata</i>
Triatoma_brasiliensisVVL24.jpg	<i>Triatoma brasiliensis</i>	<i>Triatoma melanica</i>
Triatoma_brasiliensisVVL66.jpg	<i>Triatoma brasiliensis</i>	<i>Triatoma melanica</i>
Triatoma_brasiliensis_IRD14.jpg	<i>Triatoma brasiliensis</i>	<i>Triatoma melanica</i>
TriatomineowTbrasiliensis (84).jpg	<i>Triatoma brasiliensis</i>	<i>Triatoma costalimai</i>
Triatoma_breyeri176.jpg	<i>Triatoma breyeri</i>	<i>Triatoma maculata</i>
Triatoma_carcavalloi_IRR2.jpg	<i>Triatoma carcavalloi</i>	<i>Triatoma arthurneivai</i>
Triatoma_carcavalloi_SP1_1.jpg	<i>Triatoma carcavalloi</i>	<i>Triatoma rubrovaria</i>
Triatoma_carcavalloi_SP1_10.jpg	<i>Triatoma carcavalloi</i>	<i>Triatoma pintodiasi</i>
Triatoma_carcavalloi_SP1_2.jpg	<i>Triatoma carcavalloi</i>	<i>Triatoma pintodiasi</i>
Triatoma_carcavalloi_SP1_8.jpg	<i>Triatoma carcavalloi</i>	<i>Triatoma rubrovaria</i>
Triatoma_costalimai_SP1_2.jpg	<i>Triatoma costalimai</i>	<i>Eratyrus mucronatus</i>
Triatoma_costalimai_SP1_39.jpg	<i>Triatoma costalimai</i>	<i>Panstrongylus tibiamaculatus</i>
TriatomineowTcostalimai (16).jpg	<i>Triatoma costalimai</i>	<i>Triatoma lenti</i>
TriatomineowTcostalimai (5).jpg	<i>Triatoma costalimai</i>	<i>Triatoma lenti</i>
Triatoma_delponteivVL18.jpg	<i>Triatoma delponteii</i>	<i>Triatoma platensis</i>
Triatoma_dimidiata353.jpeg	<i>Triatoma dimidiata</i>	<i>Triatoma infestans</i>
Triatoma_eratyrusiformis375.jpg	<i>Triatoma eratyrusiformis</i>	<i>Eratyrus cuspidatus</i>
Triatoma_gerstaeckerivVL1.jpg	<i>Triatoma gerstaeckeri</i>	<i>Triatoma mexicana</i>

Triatoma_gerstaeckeriVVL2.jpg	<i>Triatoma gerstaeckeri</i>	<i>Triatoma phyllosoma</i>
Triatoma_guazuVVL10.jpg	<i>Triatoma guazu</i>	<i>Triatoma vandae</i>
Triatoma_guazuVVL13.jpg	<i>Triatoma guazu</i>	<i>Triatoma williami</i>
Triatoma_guazu_IRR3.jpg	<i>Triatoma guazu</i>	<i>Triatoma rubrovaria</i>
Triatoma_infestans_IRD26.jpg	<i>Triatoma infestans</i>	<i>Triatoma pseudomaculata</i>
Triatoma_infestans_IRD46.jpg	<i>Triatoma infestans</i>	<i>Triatoma melanica</i>
Triatoma_infestans_IRD48.jpg	<i>Triatoma infestans</i>	<i>Triatoma platensis</i>
TriatomeowTinfestans (46).jpg	<i>Triatoma infestans</i>	<i>Triatoma vitticeps</i>
Triatoma_juazeirensisVVL19.jpg	<i>Triatoma juazeirensis</i>	<i>Triatoma melanica</i>
Triatoma_juazeirensisVVL20.jpg	<i>Triatoma juazeirensis</i>	<i>Triatoma melanica</i>
Triatoma_juazeirensisVVL21.jpg	<i>Triatoma juazeirensis</i>	<i>Triatoma melanica</i>
Triatoma_klugi_SP1_12.jpg	<i>Triatoma klugi</i>	<i>Triatoma sherlocki</i>
Triatoma_lenti_IRR1.jpg	<i>Triatoma lenti</i>	<i>Triatoma melanica</i>
Triatoma_lenti_SP1_8.jpg	<i>Triatoma lenti</i>	<i>Triatoma melanica</i>
TriatomeowTlenti (12).jpg	<i>Triatoma lenti</i>	<i>Triatoma costalimai</i>
Triatoma_longipennisVVL30.jpg	<i>Triatoma longipennis</i>	<i>Triatoma phyllosoma</i>
Triatoma_maculata_IRR2.jpg	<i>Triatoma maculata</i>	<i>Triatoma pseudomaculata</i>
Triatoma_maculata_SP1_37.jpg	<i>Triatoma maculata</i>	<i>Eratyrus mucronatus</i>
Triatoma_matogrossensis164.jpg	<i>Triatoma matogrossensis</i>	<i>Triatoma maculata</i>
Triatoma_matogrossensisVVL2.jpg	<i>Triatoma matogrossensis</i>	<i>Triatoma sordida</i>
Triatoma_matogrossensis_IRR1.jpg	<i>Triatoma matogrossensis</i>	<i>Triatoma vandae</i>
Triatoma_matogrossensis_SP1_14.jpg	<i>Triatoma matogrossensis</i>	<i>Triatoma vandae</i>
Triatoma_matogrossensis_SP1_3.jpg	<i>Triatoma matogrossensis</i>	<i>Panstrongylus lignarius</i>
Triatoma_matogrossensis_SP2_1.jpg	<i>Triatoma matogrossensis</i>	<i>Triatoma vandae</i>
Triatoma_mazzottiVVL18.jpg	<i>Triatoma mazzotti</i>	<i>Triatoma phyllosoma</i>
Triatoma_melanica_SP1_3.jpg	<i>Triatoma melanica</i>	<i>Triatoma brasiliensis</i>
Triatoma_petrocchiae_SP2_6.jpg	<i>Triatoma petrocchiae</i>	<i>Triatoma brasiliensis</i>
Triatoma_petrocchiae_SP2_8.jpg	<i>Triatoma petrocchiae</i>	<i>Triatoma vandae</i>
Triatoma_petrocchiae_SP3_1.jpg	<i>Triatoma petrocchiae</i>	<i>Triatoma melanica</i>
Triatoma_phyllosomaVVL1.jpg	<i>Triatoma phyllosoma</i>	<i>Triatoma mazzotti</i>
Triatoma_phyllosomaVVL10.jpg	<i>Triatoma phyllosoma</i>	<i>Triatoma mazzotti</i>
Triatoma_phyllosomaVVL11.jpg	<i>Triatoma phyllosoma</i>	<i>Triatoma mazzotti</i>
Triatoma_phyllosomaVVL52.jpg	<i>Triatoma phyllosoma</i>	<i>Triatoma pallidipennis</i>
Triatoma_pintodiasi_SP1_16.jpg	<i>Triatoma pintodiasi</i>	<i>Triatoma delponteii</i>
Triatoma_platensisVVL4.jpg	<i>Triatoma platensis</i>	<i>Triatoma delponteii</i>
Triatoma_platensisVVL5.jpg	<i>Triatoma platensis</i>	<i>Triatoma delponteii</i>
Triatoma_platensis_IRR2.jpg	<i>Triatoma platensis</i>	<i>Triatoma delponteii</i>
Triatoma_pseudomaculata_IRD49.jpg	<i>Triatoma pseudomaculata</i>	<i>Triatoma vitticeps</i>
Triatoma_pseudomaculata_IRR27.jpg	<i>Triatoma pseudomaculata</i>	<i>Triatoma maculata</i>
Triatoma_rubrovaria_SP7_20.jpg	<i>Triatoma rubrovaria</i>	<i>Triatoma pintodiasi</i>
Triatoma_sherlocki_IRD14.jpg	<i>Triatoma sherlocki</i>	<i>Triatoma melanica</i>
Triatoma_sherlocki_SP1_3.jpg	<i>Triatoma sherlocki</i>	<i>Triatoma melanica</i>
TriatomeowTsordida (20).jpg	<i>Triatoma sordida</i>	<i>Triatoma infestans</i>
Triatoma_vitticepsVVL1.jpg	<i>Triatoma vitticeps</i>	<i>Triatoma melanocephala</i>
Triatoma_vitticeps_IRD4.jpg	<i>Triatoma vitticeps</i>	<i>Panstrongylus megistus</i>
Triatoma_vitticeps_IRR7.jpg	<i>Triatoma vitticeps</i>	<i>Panstrongylus lignarius</i>
Triatoma_williami_IRR4.jpg	<i>Triatoma williami</i>	<i>Triatoma wygodzinsky</i>
Triatoma_wygodzinsky_IRR2.jpg	<i>Triatoma wygodzinsky</i>	<i>Triatoma pseudomaculata</i>

APÊNDICE 4

Apêndice 4. Comparação do desempenho do AlexNet neste estudo (VLM) com as técnicas usadas em Gurgel-Gonçalves *et al.* (2017) 'RGG' e Khalighifar *et al.* (2019) 'AK'. Os destaques coloridos mostram os desempenhos negativos comparados aos estudos anteriores.

Espécie	Acurácia específica			Diferença	
	VLM	RGG	AK	VLM-RGG	VLM-AK
<i>Cavernicola lenti</i>	100,0%	93,3%	100,0%	6,7%	0,0%
<i>Cavernicola pilosa</i>	66,7%				
<i>Dipetalogaster maxima</i>	100,0%				
<i>Eratyrus cuspidatus</i>	100,0%				
<i>Eratyrus mucronatus</i>	94,7%	80,0%		14,7%	
<i>Panstrongylus diasi</i>	100,0%	96,7%	95,2%	3,3%	4,8%
<i>Panstrongylus geniculatus</i>	100,0%	93,3%	100,0%	6,7%	0,0%
<i>Panstrongylus lignarius</i>	98,6%	85,7%	93,1%	12,9%	5,5%
<i>Panstrongylus lutzi</i>	100,0%	88,2%	88,2%	11,8%	11,8%
<i>Panstrongylus megistus</i>	100,0%	91,7%	95,2%	8,3%	4,8%
<i>Panstrongylus rufotuberculatus</i>	60,0%	100,0%	100,0%	-40,0%	-40,0%
<i>Panstrongylus tibiamaculatus</i>	100,0%	92,7%	85,0%	7,3%	15,0%
<i>Psammolestes coreodes</i>	77,8%				
<i>Psammolestes tertius</i>	91,2%	100,0%	97,1%	-8,8%	-5,9%
<i>Rhodnius brethesi</i>	100,0%	96,4%	96,3%	3,6%	3,7%
<i>Rhodnius domesticus</i>	93,3%	96,3%	96,6%	-3,0%	-3,3%
<i>Rhodnius marabaensis</i>	90,9%				
<i>Rhodnius milesi</i>	100,0%	89,2%	73,8%	10,8%	26,2%
<i>Rhodnius montenegrensis</i>	95,0%	84,6%	88,1%	10,4%	6,9%
<i>Rhodnius nasutus</i>	94,1%	82,2%	88,6%	11,9%	5,5%
<i>Rhodnius neglectus</i>	100,0%	83,3%	90,9%	16,7%	9,1%
<i>Rhodnius pictipes</i>	97,1%	95,3%	97,6%	1,8%	-0,5%
<i>Rhodnius prolixus</i>	93,3%				
<i>Rhodnius robustus</i>	85,7%				
<i>Rhodnius stali</i>	96,3%				
<i>Triatoma arthurneivai</i>	80,0%	78,1%	83,9%	1,9%	-3,9%
<i>Triatoma baratai</i>	100,0%	82,8%	89,3%	17,2%	10,7%
<i>Triatoma barberi</i>	100,0%	72,4%	80,0%	27,6%	20,0%
<i>Triatoma brasiliensis</i>	92,3%	76,6%	92,1%	15,7%	0,2%
<i>Triatoma breyeri</i>	85,7%				
<i>Triatoma carcavalloii</i>	54,5%	86,8%	89,2%	-32,3%	-34,7%
<i>Triatoma circummaculata</i>	100,0%	85,7%	95,2%	14,3%	4,8%
<i>Triatoma costalimai</i>	89,5%	85,7%	90,5%	3,8%	-1,0%
<i>Triatoma delpontei</i>	90,0%	86,7%	75,9%	3,3%	14,1%
<i>Triatoma dimidiata</i>	97,1%	76,6%	86,1%	20,5%	11,0%
<i>Triatoma eratyrusiformis</i>	66,7%				
<i>Triatoma gerstaeckeri</i>	33,3%	83,3%	83,3%	-50,0%	-50,0%

<i>Triatoma guazu</i>	57,1%	64,3%	71,4%	-7,2%	-14,3%
<i>Triatoma infestans</i>	91,8%	83,3%	87,0%	8,5%	4,8%
<i>Triatoma juazeirensis</i>	57,1%	81,0%	81,0%	-23,9%	-23,9%
<i>Triatoma jurbergi</i>	100,0%				
<i>Triatoma klugi</i>	87,5%				
<i>Triatoma lenti</i>	84,2%	78,9%	78,9%	5,3%	5,3%
<i>Triatoma longipennis</i>	90,9%	72,5%	76,9%	18,4%	14,0%
<i>Triatoma maculata</i>	93,3%	89,7%	76,9%	3,6%	16,4%
<i>Triatoma matogrossensis</i>	71,4%	75,0%	81,3%	-3,6%	-9,9%
<i>Triatoma mazzottii</i>	80,0%	77,3%	63,6%	2,7%	16,4%
<i>Triatoma melanica</i>	95,7%	79,3%	89,3%	16,4%	6,4%
<i>Triatoma melanocephala</i>	100,0%	88,0%		12,0%	
<i>Triatoma mexicana</i>	100,0%	80,0%	97,8%	20,0%	2,2%
<i>Triatoma nitida</i>	100,0%	46,7%	73,9%	53,3%	26,1%
<i>Triatoma pallidipennis</i>	100,0%	90,7%	93,2%	9,3%	6,8%
<i>Triatoma petrocchiai</i>	66,7%				
<i>Triatoma phyllosoma</i>	66,7%	46,6%	69,0%	20,1%	-2,3%
<i>Triatoma pintodiasi</i>	93,3%		84,0%		9,3%
<i>Triatoma platensis</i>	62,5%	74,1%	70,4%	-11,6%	-7,9%
<i>Triatoma pseudomaculata</i>	93,8%	70,9%	74,5%	22,9%	19,3%
<i>Triatoma rubrofaciata</i>	100,0%				
<i>Triatoma rubrovaria</i>	98,2%	59,3%	79,6%	38,9%	18,6%
<i>Triatoma sherlocki</i>	90,0%	93,5%	90,3%	-3,5%	-0,3%
<i>Triatoma sordida</i>	97,8%	81,2%	92,6%	16,6%	5,2%
<i>Triatoma vanda</i>	100,0%	69,0%	79,3%	31,0%	20,7%
<i>Triatoma vitticeps</i>	88,0%	85,1%	80,9%	2,9%	7,1%
<i>Triatoma williami</i>	85,7%	70,6%	70,0%	15,1%	15,7%
<i>Triatoma wygodzinsky</i>	83,3%				

CAPÍTULO III

Identificação automática de ninfas de triatomíneos ao nível de gênero usando a rede neural convolucional pré-treinada AlexNet

Manuscrito em elaboração

Resumo

Este estudo apresenta uma abordagem inovadora para a identificação automatizada de ninfas de triatomíneos, vetores da doença de Chagas, por meio da aplicação da rede neural convolucional AlexNet. Foram utilizadas 2.432 imagens de insetos, incluindo ninfas de triatomíneos e outras espécies heterópteras, para treinar e testar a rede em três tarefas distintas. A acurácia da AlexNet para identificação de ninfas de triatomíneos versus heterópteros não triatomíneos (adultos e ninfas) atingiu ~0,986, enquanto a identificação de ninfas de triatomíneos versus ninfas de não triatomíneos apresentou uma acurácia de ~0,991. Além disso, a acurácia da rede para identificar os gêneros mais relevantes em saúde pública (*Panstrongylus*, *Rhodnius* e *Triatoma*) foi de ~0,941. Esses resultados indicam a viabilidade da utilização da AlexNet como uma ferramenta de identificação precisa e eficaz para ninfas de triatomíneos. Essa abordagem pode ter implicações significativas para a vigilância e controle da doença de Chagas, permitindo uma identificação mais rápida e confiável dos vetores. Embora seja um primeiro passo, este estudo ressalta o potencial do aprendizado de máquina para enfrentar desafios de identificação taxonômica, fornecendo uma base sólida para futuros desenvolvimentos e melhorias. A expansão contínua do banco de imagens e a validação em cenários da vida real podem aprimorar ainda mais essa ferramenta, tornando-se um recurso fundamental na vigilância e controle da doença de Chagas e promovendo a saúde pública.

Palavras-chave: Heteroptera, ninfas, Triatominae, identificação, aprendizado de máquina.

1. Introdução

Os triatomíneos desempenham um papel importante em saúde pública, uma vez que atuam como vetores de *Trypanosoma cruzi*, que causa a doença de Chagas (WHO, 2021). A forma clássica de infecção em humanos ocorre quando o triatomíneo (adulto ou ninfa) defeca na pele, durante a hematofagia, e o parasito penetra no corpo, multiplicando-se e ocasionando lesões principalmente no coração, esôfago e intestino (Rassi *et al.*, 2010). Outras formas de transmissão descritas são por meio da ingestão de alimentos infectados, incidentes em laboratório, transfusão sanguínea, materno infantil e transplante de órgãos (WHO, 2019).

Mais de 150 espécies de triatomíneos são conhecidas no mundo, e todas podem transmitir *T. cruzi* (Barrett, 1991; Monteiro *et al.*, 2018, Zhao *et al.*, 2023). Essa grande diversidade de espécies faz com que seja necessária uma vigilância efetiva preparada para identificar corretamente as espécies que são encontradas invadindo ou colonizando as residências. Tradicionalmente a identificação de triatomíneos é feita por chaves dicotômicas, porém elas estão sempre desatualizadas e requerem muito treinamento (por exemplo, Lent & Wygodzinsky, 1979; Galvão, 2014), além de não cobrirem de forma adequada a identificação de ninfas (Carcavallo *et al.* 1997), as quais também são frequentes no ambiente domiciliar e necessitam ser identificadas.

Esforços já foram realizados para facilitar a identificação desses vetores, por exemplo, a criação de chaves eletrônicas como o TriatoKey (Oliveira *et al.*, 2017) e o TriatoDex (Gurgel-Gonçalves *et al.*, 2021). No entanto, apesar de serem interativas e dinâmicas em relação às chaves convencionais, essas ferramentas também necessitam de usuários treinados para interpretar as características morfológicas dos indivíduos. Além disso, nenhuma dessas chaves eletrônicas estão preparadas para realizar a identificação de ninfas dos triatomíneos.

Uma alternativa para esse problema é a aplicação de técnicas de aprendizado de máquina para a identificação automática das espécies a partir de imagens. Estudos anteriores mostraram que a utilização dessas técnicas pode contribuir como um grande avanço na identificação de vetores de doenças tropicais (Justen *et al.*, 2021; Lorenz *et al.*, 2015; Motta *et al.*, 2019; 2020; Park *et al.*, 2020; Pataki *et al.*, 2021; Gurgel-Gonçalves *et al.*, 2017). A partir dessas perspectivas, as técnicas de aprendizado de máquina baseadas em aprendizagem profunda se destacam (Khalighifar *et al.*, 2019; Motta *et al.*, 2019). O primeiro sistema de identificação de triatomíneos desenvolvido utilizou classificadores estatísticos para a identificação de imagens de 51 espécies, obtendo sucesso de identificação de ~80%, sendo melhorado para ~90% quando se considerava apenas espécies que ocorrem no Brasil (Gurgel-Gonçalves *et al.*, 2017). Esse sucesso de identificação aumentou quando se utilizou uma técnica de aprendizagem profunda para identificação do mesmo conjunto de imagens, alcançando até 98,9% para algumas espécies (Khalighifar *et al.*, 2019). Em outros estudos também foram alcançados bons resultados na identificação de triatomíneos com aplicação de diferentes técnicas de aprendizado de máquina (Cruz *et al.*, 2020; Abdelghani *et al.*, 2021; Parsons & Banitaan, 2021; Cochero *et al.*, 2022, capítulos 1 e 2). No entanto, nenhuma dessas abordagens consideraram em suas

análises a identificação de imagens de ninfas de triatomíneos, que são muito importantes para saúde pública, pois quando são encontradas nas residências, indicam eventos de colonização, o que pode aumentar drasticamente o risco de infecção dos humanos por *T. cruzi* Chagas, 1909 (Ribeiro Jr *et al.*, 2021).

A partir deste problema, buscamos aplicar uma técnica de aprendizado de máquina baseada em aprendizagem profunda, a rede neural convolucional pré-treinada AlexNet (Krizhevsky *et al.*, 2012), a qual foi eficaz para a identificação de imagens dorsais de triatomíneos adultos até o nível de gênero (capítulo 1). Para alcançar esta meta, executamos três experimentos buscando avaliar o desempenho da rede AlexNet na identificação de: a) ninfas de triatomíneos e hemípteros não triatomíneos (adultos e ninfas); b) ninfas de triatomíneos e ninfas de não triatomíneos; e c) ninfas de triatomíneos dos gêneros *Panstrongylus* Berg, 1879, *Rhodnius* Stål, 1859 e *Triatoma* Laporte, 1832. A nossa hipótese é que esta técnica de aprendizagem terá bom desempenho para essas tarefas estabelecidas, contribuindo para melhoria dos sistemas de vigilância de triatomíneos no futuro.

2. Material e métodos

2.1. Banco de imagens

Nós usamos uma amostra de 2.432 imagens de insetos, incluindo 1.111 de hemípteros não-triatomíneos (adultos e ninfas), 609 de ninfas de não-triatomíneos e 712 ninfas de triatomíneos. Foram fotografados os espécimes disponíveis no laboratório de Parasitologia Médica e Biologia de Vetores da Universidade de Brasília (UnB), no Laboratório de Patologia e Biologia Molecular do Instituto Gonçalo Muniz (Fiocruz-BA), Laboratório de Parasitologia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) e no Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos do Instituto Oswaldo Cruz (Fiocruz-RJ). A identificação dos triatomíneos foi confirmada utilizando chaves impressas (Lent & Wygodzinsky, 1979; Galvão & Dale., 2014), e informações sobre a origem geográfica dos espécimes. Parte das imagens foram obtidas pelo equipamento apresentado em Gurgel-Gonçalves *et al.* (2017), e por uma câmera de celular comum. Outra parte foi obtida a partir de sites especializados, considerando a identificação a nível de subordem (Heteroptera, Latreille, 1810) (apêndice 1). As imagens apresentavam características variadas, como diferentes cores de fundo, texturas e posições.

2.2. Algoritmo: treino e teste

Nós usamos a rede AlexNet, um algoritmo de aprendizagem profunda que implementa uma Rede Neural Convolutiva (CNN) que realiza o aprendizado de transferência, tomando uma rede pré-treinada como fonte de alguns recursos de entrada antes de aprender com as imagens do conjunto de dados. Para o aprendizado de máquina, configuramos a rede da seguinte forma: do total de imagens foram destinadas 70% para treinamento, 20% para teste e 10% para validação interna. As análises foram executadas no ambiente computacional MATLAB (www.mathworks.com). A AlexNet é uma rede neural convolutiva (CNN) pré-treinada com 1,2 milhão de imagens de alta resolução, classificadas em mais de 1000 categorias diferentes com 60 milhões de parâmetros (Krizhevsky *et al.*, 2012). Todas as imagens pré-processadas foram redimensionadas para 227× 227 pixels, padrões necessários para esta rede. Durante o processo de aprendizagem, o AlexNet escolhe de forma aleatória as imagens de treinamento, de teste e de validação. Ao dividir os dados dessa maneira, a avaliação do desempenho do modelo se torna independente do conjunto de treinamento, garantindo que a rede neural seja capaz de funcionar bem para dados ainda não vistos. Para o treinamento, utilizamos o otimizador "stochastic gradient descent with momentum" com as configurações padrão da rede, exceto para a taxa de aprendizado inicial (que definimos como 0.001) e o número máximo de épocas (que definimos como 30 para as tarefas de identificação de 'ninfas de triatomíneos × não triatomíneos (adultos e ninfas)' e para 'ninfas de triatomíneos × ninfas de não triatomíneos'; e 50 para a tarefa de identificação de ninfas dos gêneros, '*Panstrongylus* × *Rhodnius* × *Triatoma*' (Figura 1), após testes preliminares para observar como o desempenho da rede se comportava). Resultados binários das identificações (corretas/incorretas) foram usados como variáveis nas análises (ver Miranda *et al.*, 2023, capítulo 1).

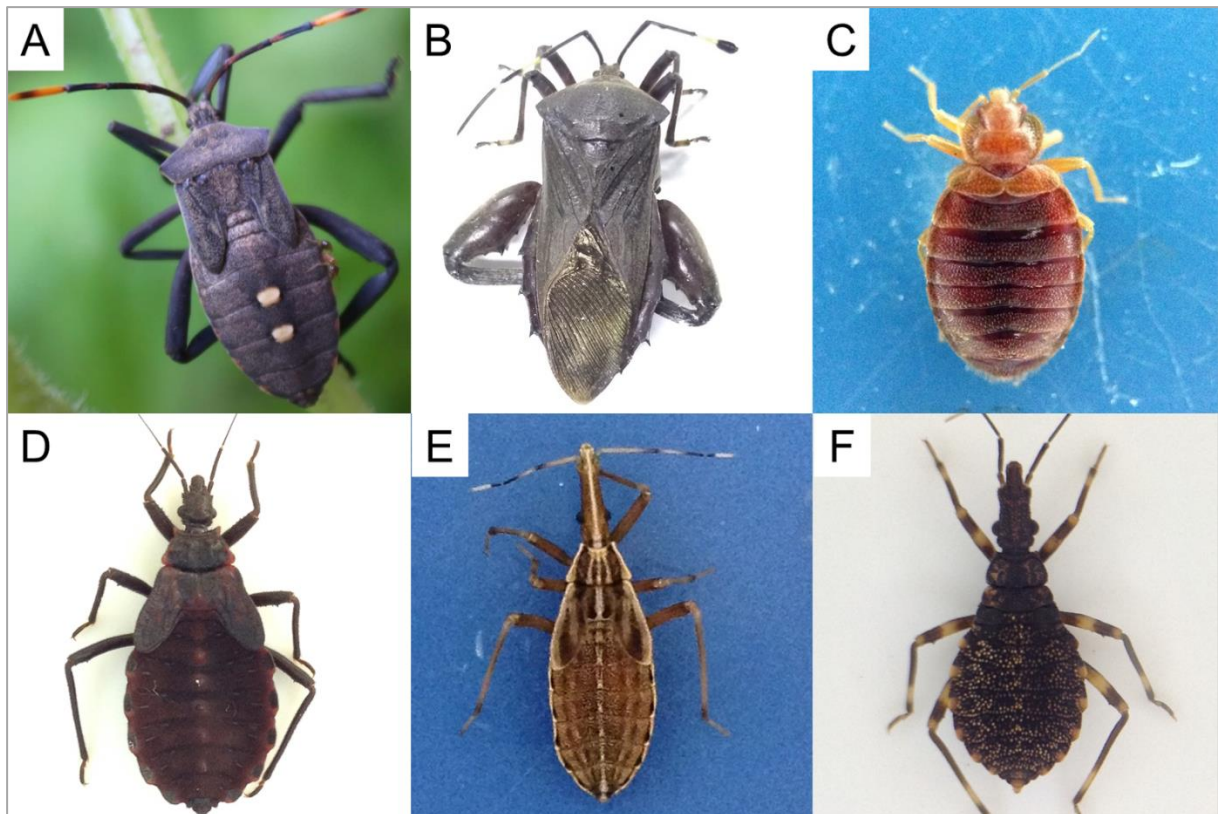


Figura 1. Exemplos de grupos de insetos que foram utilizados nos experimentos. A – Ninfa de percevejo fitófago (Coreidae Leach, 1815). B – Adulto de percevejo fitófago (Coreidae). C – Percevejo de cama (Cimicidae Latreille, 1802). D, E e F – Ninfas de triatomíneos dos gêneros *Panstrongylus*, *Rhodnius* e *Triatoma* (Reduviidae Latreille, 1807), respectivamente.

2.3. Análise de dados

Nós realizamos três experimentos com a AlexNet. No primeiro, treinamos o algoritmo para aprender a diferenciar entre ninfas de triatomíneos e espécimes de não-triatomíneos (como percevejos heterópteros que não pertencem a subfamília Triatominae Jeannel, 1919 tanto em fase de ninfa quanto em fase adulta), sendo utilizadas 444 imagens de teste. No segundo, a tarefa envolvia diferenciar ninfas de triatomíneos de ninfas de não-triatomíneo (incluindo apenas fotos de estágios de ninfas nas duas classes testadas), totalizando 344 imagens de teste. E, no terceiro, a tarefa foi diferenciar ninfas dos três gêneros de triatomíneos mais relevantes em saúde pública, *Panstrongylus*, *Rhodnius* e *Triatoma* com 136 imagens de teste (neste teste incluímos apenas imagens que tínhamos certeza sobre a identificação a nível de espécie, sendo 88 de *Panstrongylus*, 82 de *Rhodnius* e 509 de *Triatoma*). Analisamos as curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda validação para destacar tendências gerais do treinamento. Após o treino do

algoritmo, nós avaliamos o seu desempenho a partir de matrizes de confusão/heatmaps com a quantidade de erros de identificação, os resultados de acurácia geral da rede (probabilidade de acertar a identificação independente da classe), acurácia específica para cada classe (sensibilidade) e especificidade. Essas métricas foram calculadas da seguinte forma:

$$\text{Acurácia geral} = \frac{VP+VN}{VP+VN+FP+FN}$$

$$\text{Acurácia específica} = \frac{VP}{VP+FN}$$

$$\text{Especificidade} = \frac{VN}{FP+VN}$$

VP = Verdadeiros positivos

VN = Verdadeiros negativos

FP = Falsos positivos

FN = Falsos negativos

Nós apresentamos os resultados em tabelas descritivas e gráficos ilustrados e realizamos análises exploratórias simples calculando as frequências e proporções com intervalos de confiança (ICs) de 'Wilson' de 95% (Newcombe, 1998) usando o pacote R 'Hmisc' (Harrell e Dupont, 2023) e para facilitar a interpretação dos resultados, separamos as ilustrações por grupos seguindo a sequência: 'ninfas de triatomíneos x não triatomíneos', 'ninfas de triatomíneos x ninfas de não triatomíneos', '*Panstrongylus* x *Rhodnius* x *Triatoma*'. Analisamos os dados usando o software computacional R 4.2.1 (R Core Team, 2022) e a interface RStudio 2023.03.1.446 (Posit Software, 2022).

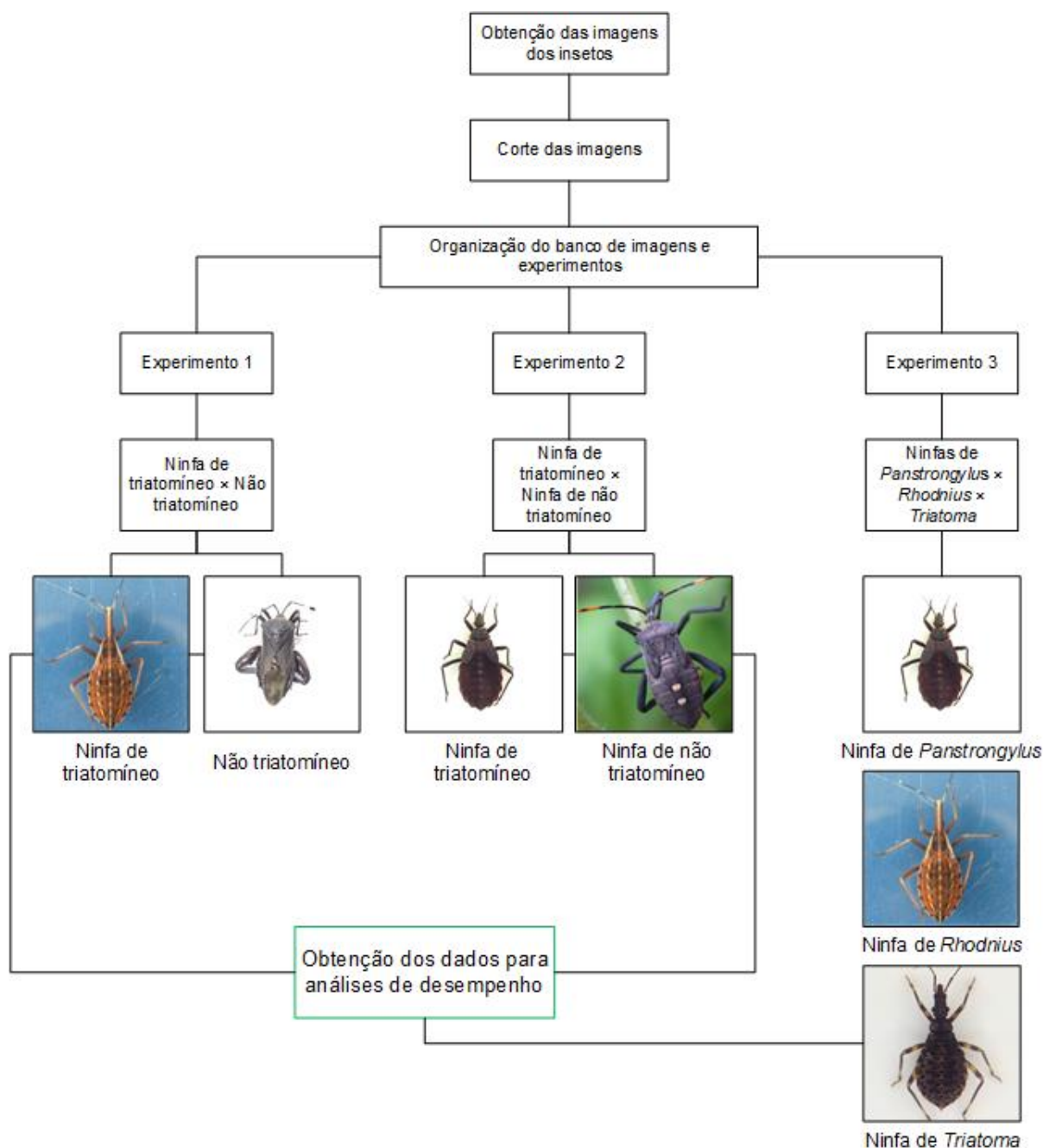


Figura 2. Fluxograma resumindo a metodologia adotada neste estudo.

3. Resultados

3.1. Resultado do treino

‘Ninfas de triatomíneos x Não triatomíneos’: A sessão de treinamento foi concluída após atingir o número máximo de épocas estabelecidas (30 épocas). Os resultados demonstraram um progresso significativo com acurácia de validação atingindo patamares altos. Ao final da sessão de treinamento, a acurácia de validação atingiu 99,10% (Figura 3).

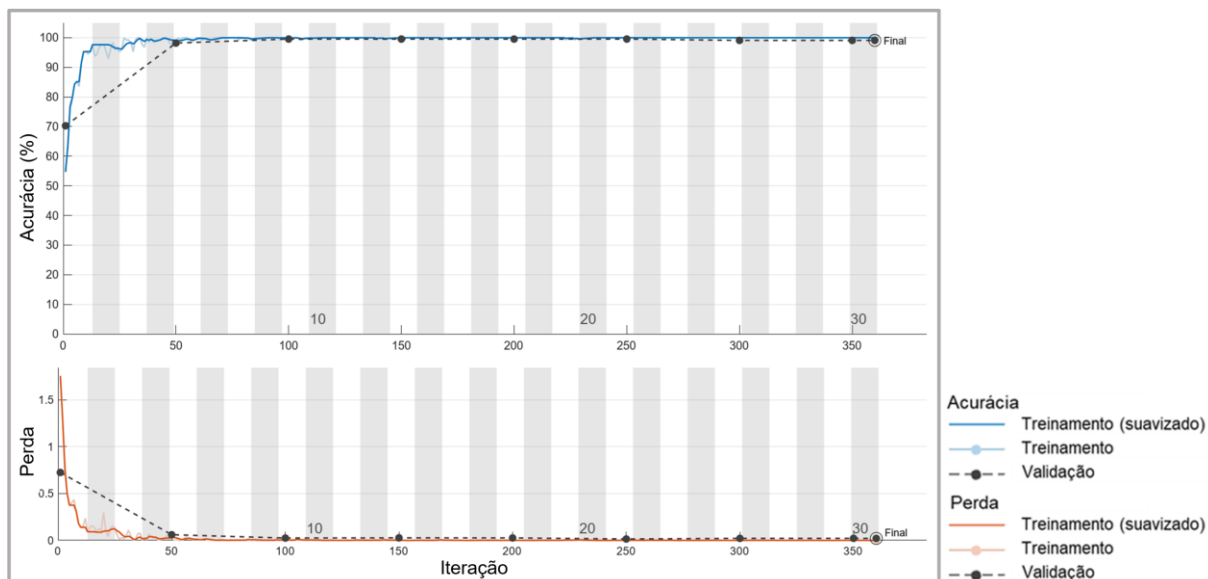


Figura 3. Progresso de treinamento da rede AlexNet. Treino incluindo duas classes, ‘ninfas de triatomíneos × não triatomíneos’. As métricas de acurácia e perda são exibidas ao longo das iterações. As curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda de validação são apresentadas, com uma suavização aplicada para destacar tendências gerais. Os eixos horizontal e vertical representam as iterações e os valores correspondentes das métricas, respectivamente.

‘Ninfas de triatomíneos × ninfas de não triatomíneos’: Os resultados também demonstraram um bom progresso de treinamento. A taxa de aprendizado permaneceu constante ao longo do treinamento, alcançando uma acurácia de validação alcançada foi de 98,26% (Figura 4).

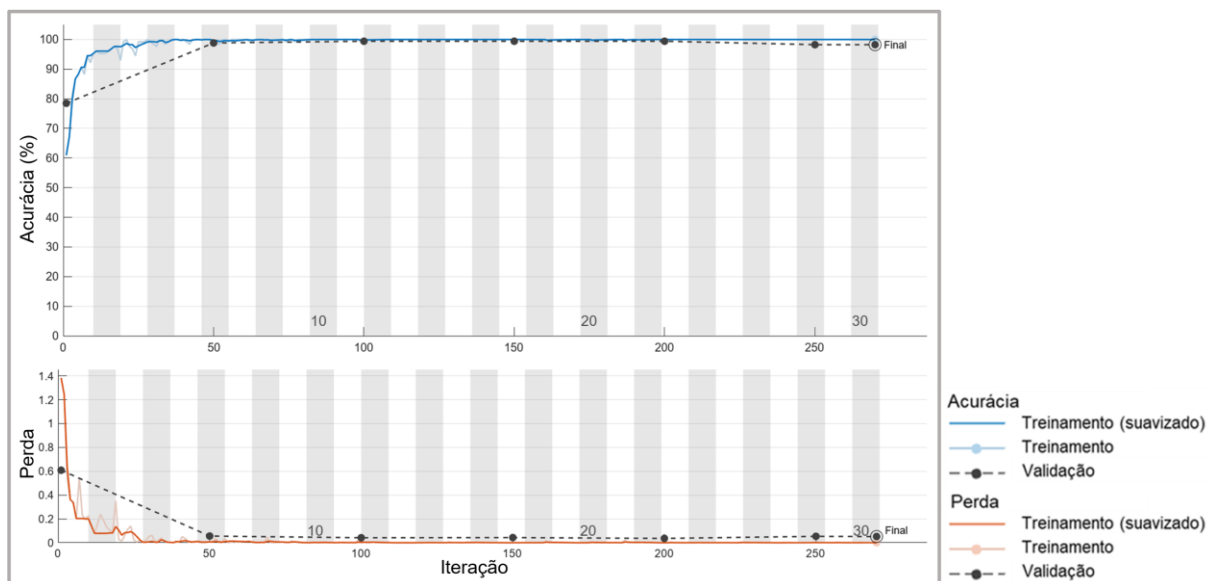


Figura 4. Progresso de treinamento da rede AlexNet. Treino incluindo duas classes, ‘ninfas de triatomíneos × ninfas de não triatomíneos’. As métricas de acurácia e perda são exibidas ao longo das iterações. As curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda de validação são apresentadas, com uma suavização aplicada para destacar tendências gerais. Os eixos horizontal e vertical representam as iterações e os valores correspondentes das métricas, respectivamente.

de treinamento e a perda de validação são apresentadas, com uma suavização aplicada para destacar tendências gerais. Os eixos horizontal e vertical representam as iterações e os valores correspondentes das métricas, respectivamente.

'*Panstrongylus x Rhodnius x Triatoma*': Neste treinamento, o número máximo de épocas foi de 50. A taxa de aprendizado também se mostrou constante até a finalização do treino. A acurácia de validação atingiu 92,65%, também indicando um desempenho promissor da rede treinado para tarefa de identificação de gêneros de triatomíneos (Figura 5).

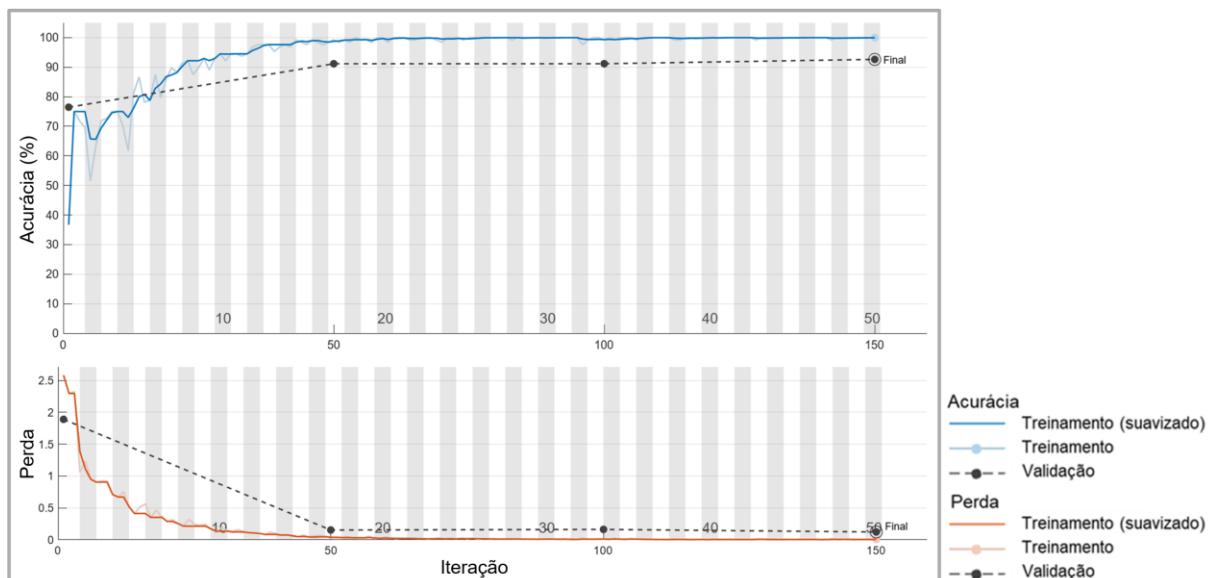


Figura 5. Progresso de treinamento da rede AlexNet. Treino incluindo três classes, '*Panstrongylus x Rhodnius x Triatoma*'. As métricas de acurácia e perda são exibidas ao longo das iterações. As curvas de acurácia de treinamento, acurácia de validação, perda de treinamento e a perda de validação são apresentadas, com uma suavização aplicada para destacar tendências gerais. Os eixos horizontal e vertical representam as iterações e os valores correspondentes das métricas, respectivamente.

3.2. Desempenho da AlexNet

Ninfa de triatomíneo x Não triatomíneo

Nesta tarefa ocorreram 438 acertos e seis erros, no qual foi alcançada uma acurácia geral de ~0,986 (IC95%, 0,970-0,993). Três ninfas de triatomíneos foram confundidas com insetos não triatomíneos (ninfas ou não de outros heterópteros), enquanto três ninfas de não triatomíneos também foram confundidas com ninfas de triatomíneos (Figura 6).

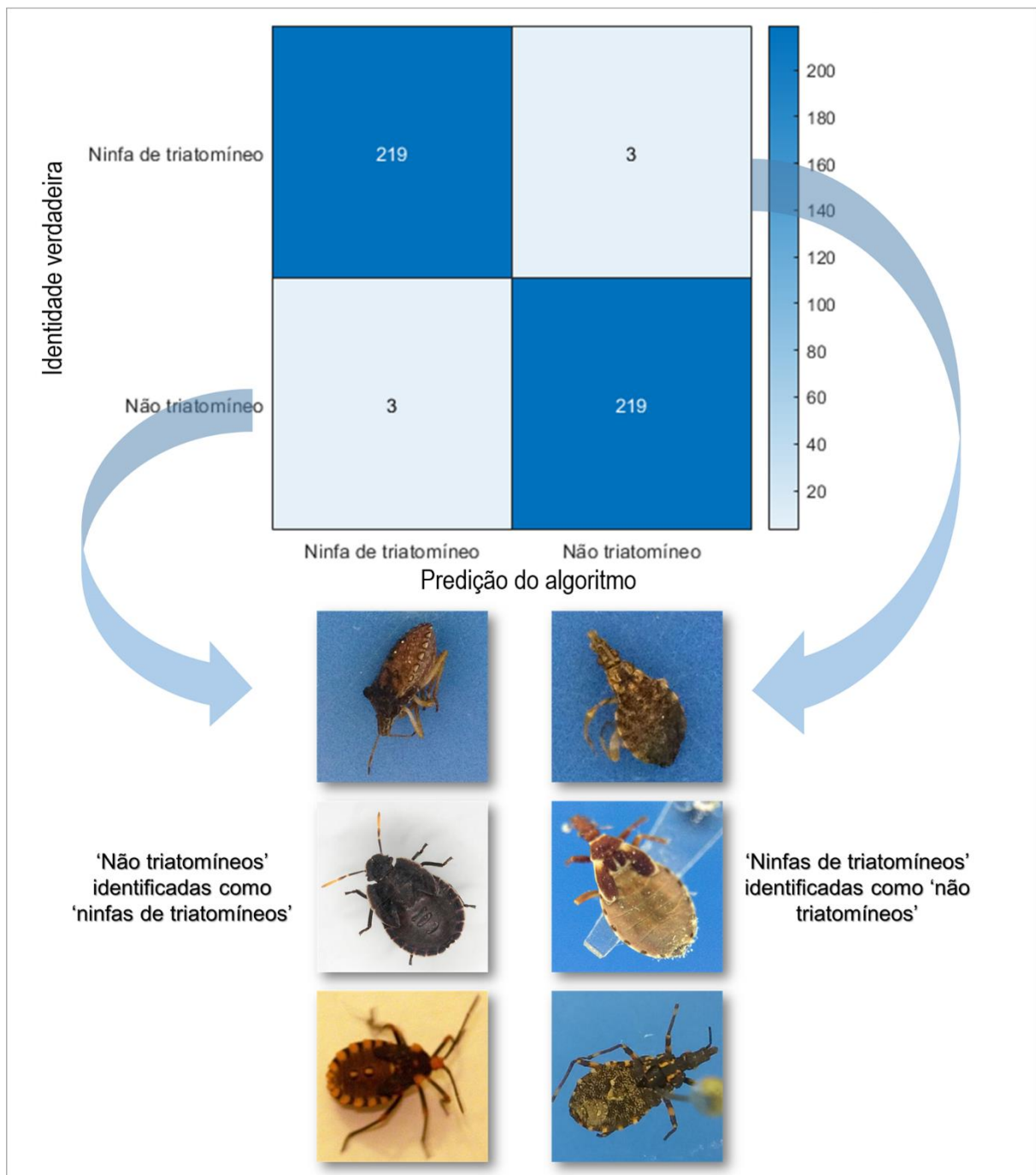


Figura 6. Representação esquemática com a matriz de confusão e os espécimes envolvidos nos erros de identificação pela rede AlexNet no experimento de identificação de 'Ninfa de triatomíneo x Não triatomíneo'. Observe que apesar de estarem incluídos na classe 'não triatomíneo', os insetos confundidos estão na fase de ninfa.

Ninfa de triatomíneo x Ninfa de não triatomíneo

Nesta tarefa ocorreram 341 acertos e três erros, com rede AlexNet alcançando uma acurácia geral de ~0,991 (IC95%, 0,974-0,997). Três 'ninfas de não

triatomíneos' foram confundidas com 'ninfas de triatomíneos', enquanto todas as 'ninfas de triatomíneos' foram identificadas corretamente (Figura 7).

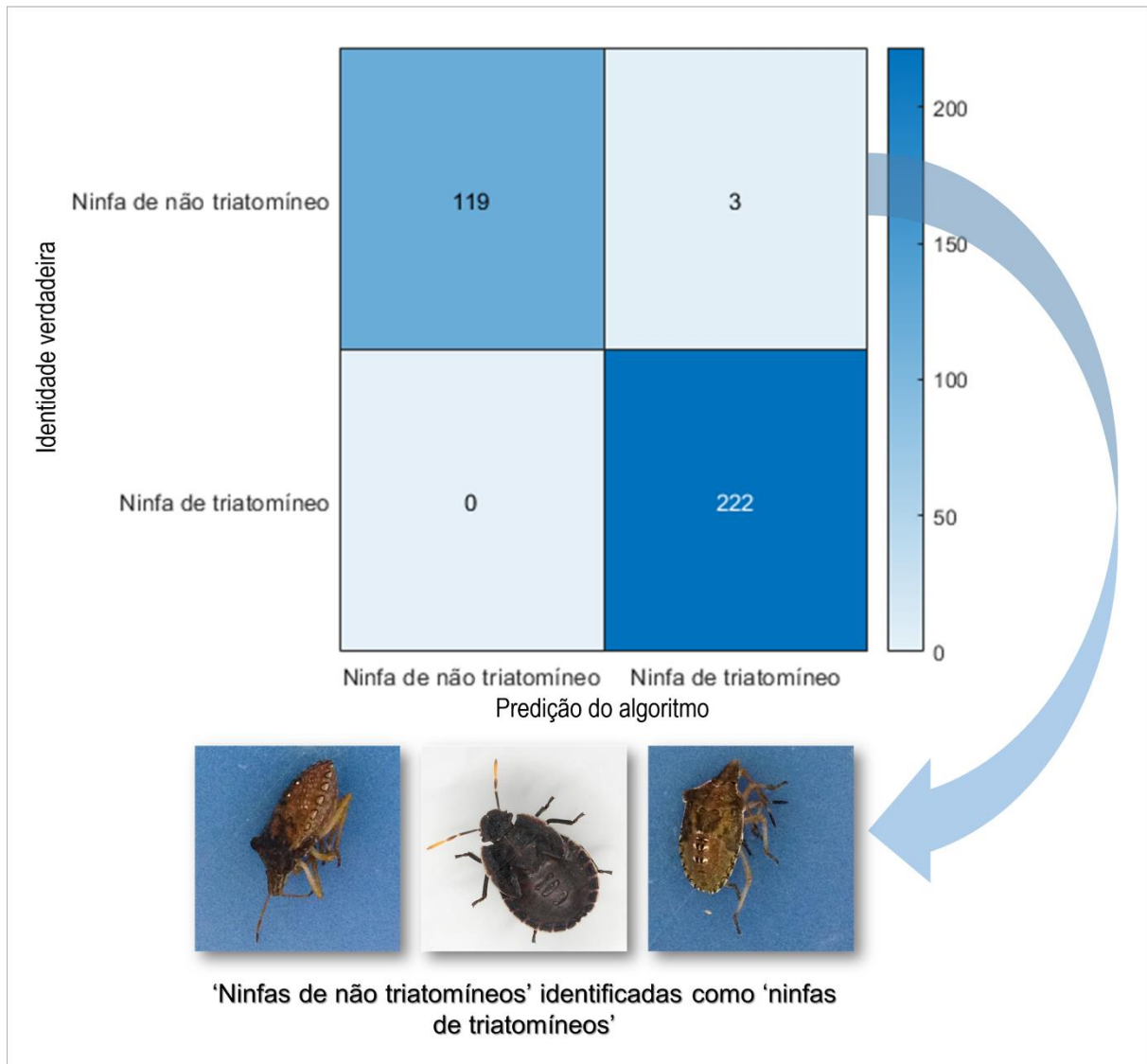


Figura 7. Representação esquemática com a matriz de confusão e os espécimes envolvidos nos erros de identificação pela rede AlexNet no experimento de identificação de 'Ninfa de triatomíneo x ninfas de não triatomíneo'. Observe que os dois espécimes da esquerda estavam envolvidos nos erros do experimento anterior.

Ninfa de *Panstrongylus* x *Rhodnius* x *Triatoma*

Nesta tarefa ocorreram 128 acertos e oito erros, com rede AlexNet alcançando uma acurácia geral de ~0,941 (IC95%, 0,888-0,969). Sete ninfas do gênero '*Panstrongylus*' foram confundidas com 'ninfas do gênero '*Triatoma*' e uma ninfa de '*Triatoma*' foi confundida com ninfa do gênero '*Rhodnius*' (Figura 8).

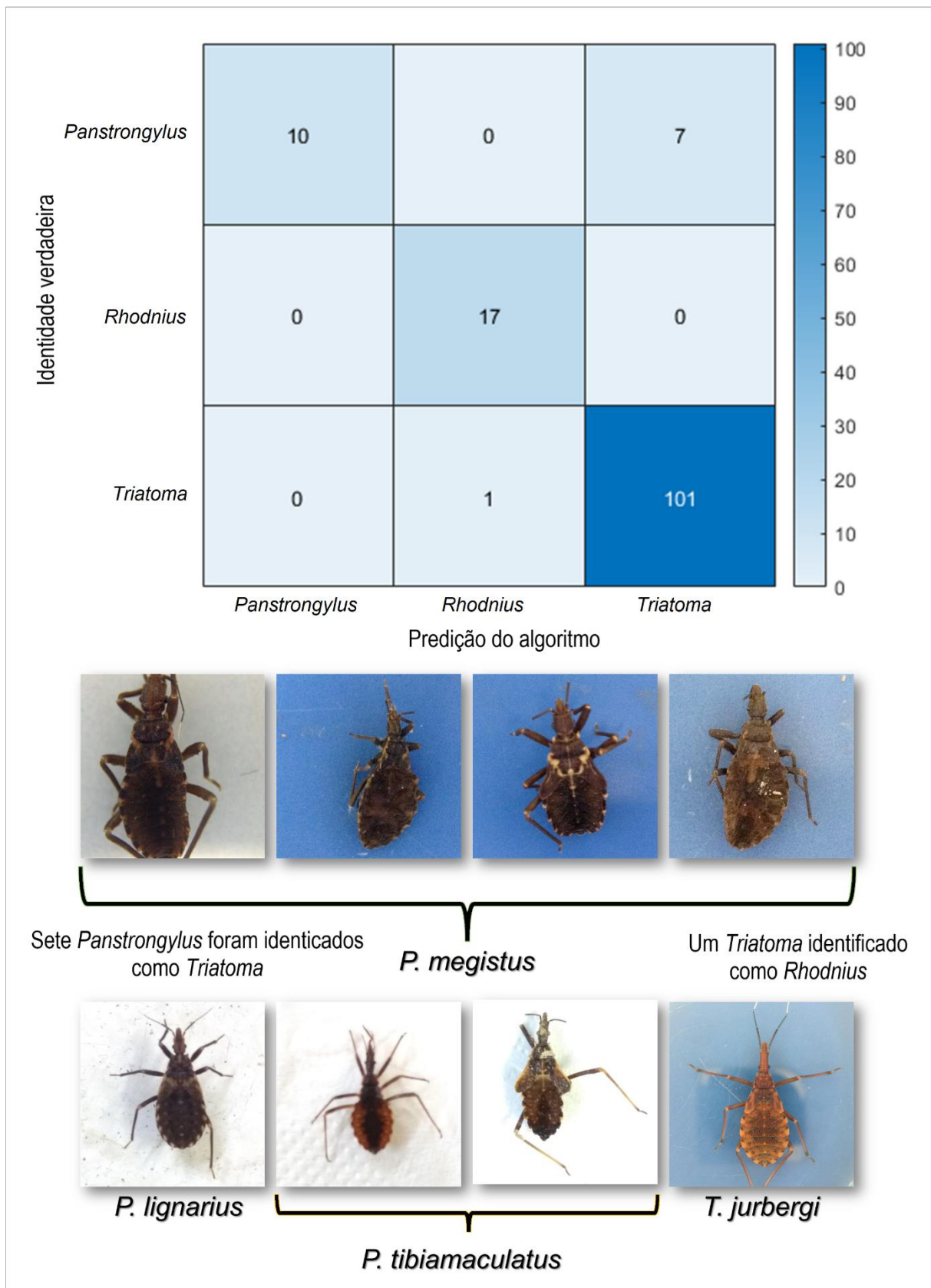


Figura 8. Representação esquemática com a matriz de confusão e os espécimes envolvidos nos erros de identificação pela rede AlexNet no experimento de identificação de ‘Ninfas de *Panstrongylus* × *Rhodnius* × *Triatoma*’. Sete ninfas do gênero *Panstrongylus* [*P. megistus*

(Burmeister, 1835), *P. lignarius* (Walker, 1873) e *P. tibiamaculatus* (Pinto, 1926)] foram identificadas como *Triatoma*. Uma ninfa do gênero *Triatoma* (*T. jurbergi* Carcavallo et al., 1998) foi identificada como *Rhodnius*.

Dos três experimentos realizados, aquele que obteve melhor acurácia geral foi o de', 'ninfas de triatomíneos × ninfas de não triatomíneos', seguido de 'ninfas de triatomíneos × não triatomíneos, e por fim, 'ninfas de *Panstrongylus* × *Rhodnius* × *Triatoma* (Figura 9).

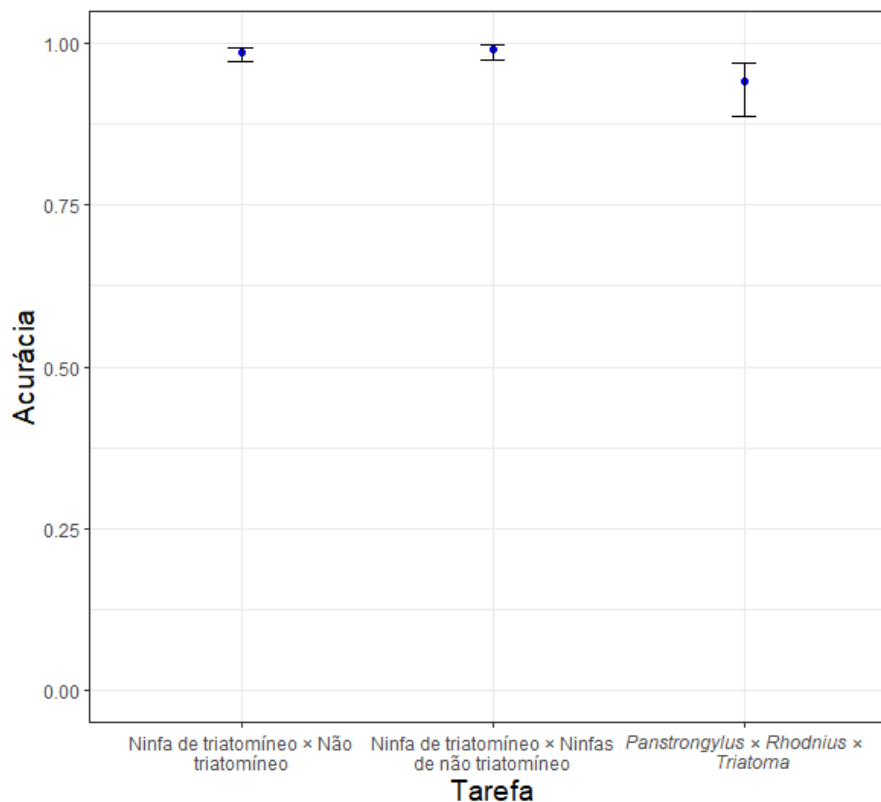


Figura 9. Acurácia geral para as três tarefas estabelecidas. As barras representam os limites do intervalo de confiança de 95%.

As probabilidades individuais, como a acurácia específica (sensibilidade) e a especificidade, mostraram o mesmo desempenho para os resultados do experimento 'ninfas de triatomíneos × não triatomíneos' apresentando acurácia específica e especificidade de ~0,986. Já para o experimento de 'ninfa de triatomíneo × ninfa de não triatomíneo', a classe ninfa de triatomíneo obteve acurácia específica máxima, e conseqüentemente, menor especificidade, já que ocorreram erros na outra classe comparada. E por fim, no experimento de ninfas de '*Panstrongylus* × *Rhodnius* × *Triatoma*', *Rhodnius* foi o gênero mais bem identificado com acurácia máxima.

Triatoma também obteve uma boa acurácia de identificação, com apenas um erro. No entanto, *Panstrongylus* apresentou sete identificações incorretas, ficando com acurácia específica de ~0,588 e especificidade de ~0,999, já que foram registrados apenas um erro de identificação para as outras duas classes comparadas (Tabela 1).

Tabela 1. Acurácia específica e especificidade observada para as três tarefas de identificação de imagens de triatomíneos e não triatomíneos (adultos e/ou ninfas) (com limites de intervalos de confiança de 95%).

Tarefa	Classe	Acurácia específica			Especificidade		
		Proporção	IC inferior	IC superior	Proporção	IC inferior	IC superior
Ninfa de triatomíneo × Não triatomíneo	Ninfa de triatomíneo	0,986	0,961	0,995	0,986	0,961	0,995
	Não triatomíneo	0,986	0,961	0,995	0,986	0,961	0,995
Ninfa de triatomíneo × Ninfa de não triatomíneo	Ninfa de triatomíneo	1,000	0,983	1,000	0,975	0,930	0,992
	Ninfa de não triatomíneo	0,975	0,930	0,992	1,000	0,983	1,000
<i>Panstrongylus</i> × <i>Rhodnius</i> × <i>Triatoma</i>	<i>Panstrongylus</i>	0,588	0,360	0,784	0,992	0,954	0,999
	<i>Rhodnius</i>	1,000	0,816	1,000	0,933	0,873	0,966
	<i>Triatoma</i>	0,990	0,947	0,998	0,794	0,632	0,897

4. Discussão

Nossos resultados mostraram que a rede neural convolucional AlexNet apresentou um ótimo desempenho para identificação de ninfas dos vetores da doença de Chagas, alcançando acurácias de até ~0,986 para identificar se a imagem era de ninfa de triatomíneo ou não. Nessa tarefa a AlexNet confrontou imagens de ninfas de triatomíneos com imagens de outros insetos que não pertenciam a subfamília Triatominae, mas que estão classificados dentro da subordem Heteroptera, incluindo indivíduos adultos e ninfas (por exemplo, Coreidae, Pentatomidae Leach, 1815 e Cimicidae). Essa probabilidade de acerto na identificação foi levemente melhorada quando executamos um experimento considerando apenas ninfas, incluindo uma classe ninfas de triatomíneos e outra com ninfas de heterópteros que não eram da subfamília Triatominae, alcançando uma acurácia de ~0,991. Quando filtramos as imagens para um experimento que considerou apenas ninfas dos três gêneros de triatomíneos mais relevantes para a saúde pública, a acurácia foi diminuída para ~0,941, um comportamento natural, visto que na medida em estabelecemos tarefas com níveis taxonômicos mais específicos (Valan *et al.*, 2019), a rede pode apresentar dificuldade em extrair recursos suficientes para uma correta discriminação dos grupos, devido a uma maior semelhança morfológica. Entretanto, estes são resultados fundamentais para a estruturação de uma ferramenta capaz de identificar formas imaturas de triatomíneos, talvez com uma expertise melhor que a de um próprio taxonomista, levando em conta o tempo de análise, identificação e as chances de obter uma identificação correta do grupo.

Atualmente não existe nenhum estudo que tenha considerado ninfas de triatomíneos em análises de aprendizado de máquina. Os trabalhos que tiveram essa preocupação até este momento, estão relacionados a identificação de insetos praga, usando algoritmos como o suporte de vetor de máquina (SVM), o classificador Adaboost, o YOLOv3 e Tiny-YOLOv3 na identificação de *Tagosodes orizicolus* (Muir, 1926) (Delphacidae Leach, 1815) e *Tessaratomia papilhosa* (Drury, 1770) (Tessaratomidae Stål, 1864) (ver Qing *et al.*, 2017; Chen *et al.*, 2021). Nosso trabalho amplia a visão do uso de técnicas de aprendizado de máquina na identificação de triatomíneos, uma vez que, os últimos estudos usando esta subfamília como modelo, consideraram apenas espécimes adultos em suas análises (Gurgel-Gonçalves *et al.*, 2017; Khalighifar *et al.*, 2019; Cruz *et al.*, 2020; Abdelghani *et al.*, 2021; Parsons & Banitaan, 2021; Cochero *et al.* 2022, Capítulos 1 e 2).

Com a grande abundância e diversidade de espécies de percevejos existentes, existe a necessidade de se obter um sistema treinado adequadamente para identificar ninfas de triatomíneos e distingui-las de outros insetos morfológicamente similares, como por exemplo, os representantes das demais subfamílias de Reduviidae (Harpactorinae Amyot & Serville, 1843, Peiratinae Amyot & Serville, 1843, Ectrichodiinae Amyot & Serville, 1843, Saicinae Stål, 1859, Reduviinae Latreille, 1807) (Weirauch *et al.*, 2014). Isso só é possível a partir do acesso a exemplares conservados e depositados em coleções didáticas e científicas, para permitir a obtenção de fotografias e criar um banco de dados mais robusto, diverso e rotulagens taxonômicas confiáveis. Em nosso estudo por exemplo, a confirmação taxonômica a nível de espécie foi aplicada apenas para as imagens de ninfas nas quais foram obtidas a partir da visita em coleções didáticas e científicas. Em outras imagens incluídas consideramos apenas níveis taxonômicos superiores como ordem e subordem, para organização das classes dos experimentos.

A identificação das ninfas de triatomíneos tem grande relevância para a vigilância e controle da doença de Chagas, uma vez que indica a colonização das habitações o que pode aumentar o potencial de transmissão de *T. cruzi* (Ribeiro Jr. *et al.* 2021). No entanto, devido ao seu tamanho reduzido e à falta de características morfológicas distintas, a detecção das ninfas pela população enfrenta desafios relevantes nesse contexto. Por exemplo, em 2006 Silva *et al.* (2012) coletaram em Pernambuco, Brasil, 8.273 triatomíneos dos quais apenas 14% eram ninfas. Em contraste, Grijalva *et al.* (2017) examinaram triatomíneos em comunidades no Equador, encontrando ninfas em 77% dos domicílios. *R. ecuadoriensis* Lent & Leon, 1958, *P. rufotuberculatus* Champion, 1899 e *P. howardi* Neiva, 1911 foram as espécies identificadas. Isso sugere a possibilidade de subnotificação de ninfas em residências, possivelmente devido à dificuldade de reconhecimento tanto pela população quanto pelos profissionais de saúde envolvidos na vigilância. Nossos resultados indicam que a identificação automática de ninfas de triatomíneos é precisa ao utilizar a AlexNet. Isso aponta para a possibilidade de integrar o algoritmo a um aplicativo de uso generalizado na vigilância dos vetores da doença de Chagas. Isso ajudaria na detecção e notificação de ninfas, contribuindo para a eliminação de focos por meio da aplicação de inseticidas residuais.

Nosso estudo revelou que as ninfas de diferentes gêneros de triatomíneos podem ser corretamente identificadas pela rede AlexNet. Escolhemos focar nos

gêneros de maior importância epidemiológica no Brasil: *Triatoma*, *Panstrongylus* e *Rhodnius*, com o objetivo de avaliar a precisão do algoritmo. A diferenciação entre esses gêneros é fundamental devido às suas biologias e habitats. O gênero *Triatoma*, o mais relevante nesse contexto, possui uma ampla distribuição e engloba a maioria das espécies importantes para a transmissão da doença, como é o caso de *T. infestans*, que habitava apenas ambientes domiciliares no Brasil (Galvão, 2014). Por outro lado, as espécies do gênero *Panstrongylus* são mais comumente encontradas em ambientes silvestres, frequentando buracos construídos por animais, como os feitos por tatus, ou surgindo naturalmente entre as raízes de árvores. *P. megistus*, por exemplo, coloniza refúgios de roedores, ninhos de aves, ocos de árvores, galinheiros e até estruturas dentro de domicílios (Carcavallo *et al.*, 1997). Já as espécies pertencentes ao gênero *Rhodnius* podem ser encontradas em domicílios, mas são mais comuns em palmeiras (Lent & Wygodzinsky *et al.*, 1979, Abad-Franch *et al.* 2015), sendo rara a presença de ninfas dentro das casas (Silva *et al.*, 2012). Ao avaliar os resultados, observamos que a acurácia foi mais consistente para as ninfas de *Rhodnius* e *Triatoma*, considerando tanto a média quanto o intervalo de confiança. Dado que as ninfas desse gênero são as mais frequentemente encontradas em ambientes domiciliares (Silva *et al.*, 2012, Ribeiro Jr *et al.*, 2019), tais resultados são bastante promissores.

Aqui demos o primeiro passo para a construção de um banco de imagens de ninfas de triatomíneos que também poderá contribuir com a aplicação de outras técnicas mais sofisticadas que poderão surgir no futuro, para colaborar com a vigilância dos vetores da doença de Chagas e até mesmo como ferramentas de taxonomia integrativa para outros grupos de insetos. Apesar do nosso banco de imagens de ninfas ainda estar moderadamente desbalanceado, continuaremos a obter mais imagens que represente o máximo possível a diversidade de triatomíneos e insetos similares, para desenvolver um sistema cada vez mais acurado no reconhecimento dos vetores da doença de Chagas.

5. Conclusão

O presente estudo empregou com sucesso a rede neural convolucional AlexNet como uma ferramenta promissora para a identificação automática de ninfas de triatomíneos, vetores da doença de Chagas. Através de uma abordagem inovadora, alcançamos resultados notáveis, demonstrando boas acurácias em

diferentes tarefas de identificação. A acurácia atingiu aproximadamente 0,986 ao diferenciar ninfas de triatomíneos de outros insetos heterópteros, um valor que aumentou para cerca de 0,991 quando focamos exclusivamente em ninfas. Além disso, expandimos nosso estudo para incluir os gêneros mais relevantes em saúde pública - *Panstrongylus*, *Rhodnius* e *Triatoma* - obtendo uma acurácia de cerca de 0,941 para essa tarefa mais desafiadora.

A abordagem inovadora apresentada neste estudo tem implicações profundas para a vigilância e controle da doença de Chagas. A identificação de ninfas usando chaves dicotômicas é uma tarefa complexa sujeita a erros devido à morfologia intrincada desses insetos. Ao aplicar técnicas de aprendizado de máquina, especialmente a rede AlexNet, conseguimos superar essas limitações, permitindo uma identificação mais rápida, precisa e eficiente das ninfas de triatomíneos. A utilização dessa abordagem pode revolucionar a forma como lidamos com a detecção e o controle desses vetores de doenças.

No entanto, reconhecemos que este estudo é um primeiro passo e que há desafios a serem superados. A expansão contínua do nosso banco de imagens, incluindo uma maior diversidade de espécies e estágios de desenvolvimento, é fundamental para aprimorar ainda mais a precisão da identificação automatizada. Além disso, a validação de nosso sistema em condições de campo e a incorporação de outras técnicas de aprendizado de máquina podem contribuir para a criação de uma ferramenta ainda mais robusta e eficaz. Dessa forma, nossos resultados demonstram claramente o potencial da rede AlexNet para a identificação de ninfas de triatomíneos, fornecendo uma solução promissora para um desafio de grande relevância epidemiológica. À medida que continuamos a aprimorar essa abordagem, esperamos que ela se torne uma ferramenta essencial para a vigilância e controle dos vetores da doença de Chagas, contribuindo significativamente para a saúde pública e o bem-estar das comunidades afetadas.

6. Referências

- Abad-Franch, F., Lima, M. M., Sarquis, O., Gurgel-Gonçalves, R., Sánchez-Martín, M., Calzada, J., Saldaña, A., Monteiro, F. A., Palomeque, F. S., Santos, W. S., Ângulo, V. M., Esteban, L., Dias, F. B. S., Diotaiuti, L., Bar, M. E., & Gottdenker, N. L. (2015). On palms, bugs, and Chagas disease in the Americas. *Acta tropica*, 151, 126-141.
- Abdelghani, B. A., Banitaan, S., Maleki, M., & Mazen, A. (2021). Kissing bugs identification using convolutional neural network. *IEEE Access*, 9, 140539-140548.
- Barrett T. V. (1991). Advances in triatomine bug ecology in relation to Chagas disease. In Harris K.H (Org.) *Advances in disease vector research*. New York: Springer-Verlag.
- Carcavallo, R. U., Galindez, I., José, J., & Herman, L. (1997). *Atlas dos vetores da doença de Chagas nas Américas*. Rio de Janeiro: Fiocruz Editorial.
- Chen, C. J., Huang, Y. Y., Li, Y. S., Chen, Y. C., Chang, C. Y., & Huang, Y. M. (2021). Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying. *IEEE Access*, 9, 21986-21997.
- Cochero, J., Patteri, L., Balsalobre, A., Ceccarelli, S., & Marti, G. (2022). A convolutional neural network to recognize Chagas disease vectors using mobile phone images. *Ecological Informatics*, 68, 101587.
- Cruz, D. D., Arellano, E., Denis Ávila, D., & Ibarra-Cerdeña, C. N. (2020). Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: a case for the cryptic dimidiata complex. *Parasites & vectors*, 13, 1-12.
- Galvão, C. (2014). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia. Curitiba: Sociedade Brasileira de Zoologia.
- Galvão, C., & Dale, C. (2014). Chaves de identificação para adultos. In Galvão, C. (Org.). *Vetores da doença de Chagas no Brasil*. Curitiba: Sociedade Brasileira de Zoologia.
- Grijalva, M. J., Villacís, A. G., Moncayo, A. L., Ocaña-Mayorga, S., Yumiseva, C. A., & Baus, E. G. (2017). Distribution of triatomine species in domestic and peridomestic environments in central coastal Ecuador. *PLoS neglected tropical diseases*, 11(10), e0005970.

- Gurgel-Gonçalves, R., Komp, E., Campbell, L. P., Khalighifar, A., Mellenbruch, J., Mendonça, V. J., Owens, H. L., Felix, K. L. C., Peterson, A. T., & Ramsey, J. M. (2017). Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the Virtual Vector Lab. *PeerJ*, 5, e3040.
- Gurgel-Gonçalves, R., Abad-Franch, F., de Almeida, M. R., Obara, M. T., de Souza, R. D. C. M., Batista, J. A. D. S., & Rocha, D. D. A. (2021). TriatoDex, an electronic identification key to the Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease: Development, description, and performance. *Plos one*, 16(4), e0248628.
- Harrell, F., & Dupont, C. (2023). Package 'Hmisc': Harrell miscellaneous. Disponível em <http://www.who.int/mediacentre/factsheets/fs340/en/>. Acesso em: 19 jul. 2023.
- Justen, L., Carlsmith, D., Paskewitz, S. M., Bartholomay, L. C., & Bron, G. M. (2021). Identification of public submitted tick images: A neural network approach. *Plos one*, 16(12), e0260622.
- Khalighifar, A., Komp, E., Ramsey, J. M., Gurgel-Gonçalves, R., & Peterson, A. T. (2019). Deep learning algorithms improve automated identification of Chagas disease vectors. *Journal of medical entomology*, 56(5), 1404-1410.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lent, H., & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bulletin of the American museum of Natural History*, 163(3), 123-520.
- Lorenz, C., Ferraudo, A. S., & Suesdek, L. (2015). Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Tropica*, 152, 165-169.
- Monteiro, F. A., Weirauch, C., Felix, M., Lazoski, C., & Abad-Franch, F. (2018). Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. *Advances in parasitology*, 99, 265-344.
- Motta, D., Santos, A. Á. B., Winkler, I., Machado, B. A. S., Pereira, D. A. D. I., Cavalcanti, A. M., Fonseca, E. O. L., Kirchner, F., & Badaró, R. (2019). Application

- of convolutional neural networks for classification of adult mosquitoes in the field. *PloS one*, 14(1), e0210829.
- Motta, D., Santos, A. Á. B., Machado, B. A. S., Ribeiro-Filho, O. G. V., Camargo, L. O. A., Valdenegro-Toro, M. A., Kirchner, F., & Badaró, R. (2020). Optimization of convolutional neural network hyperparameters for automatic classification of adult mosquitoes. *Plos one*, 15(7), e0234959.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17(8), 857-872.
- Oliveira, L. M., Brito, R. N., Guimarães, P. A. S., Santos, R. V. M. A., Diotaiuti, L. G., Souza, R. C. M., & Ruiz, J. C. (2017). TriatoKey: a web and mobile tool for biodiversity identification of Brazilian triatomine species. *Database*, 2017, bax033.
- Park, J., Kim, D. I., Choi, B., Kang, W., & Kwon, H. W. (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Scientific reports*, 10(1), 1012.
- Parsons, Z., & Banitaan, S. (2021). Automatic identification of Chagas disease vectors using data mining and deep learning techniques. *Ecological Informatics*, 62, 101270.
- Pataki, B. A., Garriga, J., Eritja, R., Palmer, J. R., Bartumeus, F., & Csabai, I. (2021). Deep learning identification for citizen science surveillance of tiger mosquitoes. *Scientific reports*, 11(1), 4718.
- Posit Software. (2022). RStudio 2023.03.1.446. <https://posit.co/products/open-source/rstudio/>.
- Qing, Y. A. O., CHEN, G. T., Zheng, W. A. N. G., Zhang, C., YANG, B. J., & Jian, T. A. N. G. (2017). Automated detection and identification of white-backed planthoppers in paddy fields using image processing. *Journal of Integrative Agriculture*, 16(7), 1547-1557.
- Rassi, A., & Marin-Neto, J. A. (2010). Chagas disease. *The Lancet*, 375(9723), 1388-1402.
- R Core Team. (2022). R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Ribeiro, G., Dos Santos, C. G., Lanza, F., Reis, J., Vaccarezza, F., Diniz, C., ... & Dos Reis, M. G. (2019). Wide distribution of *Trypanosoma cruzi*-infected triatomines in the State of Bahia, Brazil. *Parasites & vectors*, 12, 1-10.
- Ribeiro-Jr, G., Abad-Franch, F., de Sousa, O. M., Dos Santos, C. G., Fonseca, E. O., Dos Santos, R. F., ... & Reis, M. G. (2021). TriatoScore: an entomological-risk score for Chagas disease vector control-surveillance. *Parasites & Vectors*, 14, 1-15.
- Silva, M. B. A., Barreto, A. V. M. S., Silva, H. A. D., Galvão, C., Rocha, D., Jurberg, J., & Gurgel-Gonçalves, R. (2012). Synanthropic triatomines (Hemiptera, Reduviidae) in the state of Pernambuco, Brazil: geographical distribution and natural *Trypanosoma* infection rates between 2006 and 2007. *Revista da Sociedade Brasileira de Medicina Tropical*, 45, 60-65.
- Valan, M., Makonyi, K., Maki, A., Vondráček, D., & Ronquist, F. (2019). Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology*, 68(6), 876-895.
- Weirauch, C., Bérenger, J. M., Berniker, L., Forero, D., Forthman, M., Frankenberg, S., Freedman, A., Gordon, E., Hoey-Chamberlain, R., Hwang, W. S., Marshall, S. A., Michael, A., Paiero, S. M., Udah, O., Watson, C., Yeo, M., Zhang, G., & Zhang, J. (2014). An illustrated identification key to assassin bug subfamilies and tribes (Hemiptera: Reduviidae). *Canadian Journal of Arthropod Identification*, 26(2), 1-115.
- WHO - World Health Organization. Chagas disease (American trypanosomiasis). (2019). Disponível em <http://www.who.int/mediacentre/factsheets/fs340/en/>. Acesso em: 11 ago. 2020.
- WHO – World Health Organization. (2021). Chagas disease (American trypanosomiasis). Disponível em https://www.who.int/health-topics/chagasdisease#tab=tab_1.
- Zhao, Y., Fan, M., Li, H., & Cai, W. (2023). Review of Kissing Bugs (Hemiptera: Reduviidae: Triatominae) from China with Descriptions of Two New Species. *Insects*, 14(5), 450.

Weirauch, C., Bérenger, J. M., Berniker, L., Forero, D., Forthman, M., Frankenberg, S., Freedman, A., Gordon, E., Hoey-Chamberlain, R., Hwang, W. S., Marshall, S. A., Michael, A., Paiero, S. M., Uдах, O., Watson, C., Yeo, M., Zhang, G., & Zhang, J. (2014). An illustrated identification key to assassin bug subfamilies and tribes (Hemiptera: Reduviidae). *Canadian Journal of Arthropod Identification*, 26(2), 1-115.

APÊNDICES

Apêndice 1. Links de sites especializados usados para obtenção de imagens de heterópteros (ninfas e adultos).

https://www.flickr.com/photos/maxwel_rocha/8190906846
https://www.flickr.com/photos/brutamonte/6421015515
https://www.flickr.com/photos/brutamonte/6981784422
http://macrobrasil.blogspot.com/2013/04/percevejos.html#more
http://macrobrasil.blogspot.com/2013/04/percevejos.html#more
https://naturezatteraquea.wordpress.com/tag/hemiptera/
https://insetosaquaticosbioindicadoress.wordpress.com/2013/05/23/ordem-hemiptera/
https://www.insetologia.com.br/2015/07/ninfa-de-percevejo-em-sao-paulo.html
https://www.insetologia.com.br/2014/07/ninfa-de-percevejo-em-sao-paulo.html
https://br.pinterest.com/pin/384354149439584985/
https://www.flickr.com/photos/brutamonte/7375728700
https://www.insetologia.com.br/2017/12/ninfas-de-hemipteros-em-goias-e-minas.html
https://commons.wikimedia.org/wiki/File:Colorful_heteroptera_nymph.jpg
https://commons.wikimedia.org/wiki/File:Some_Heteroptera_Nymph_(Lychee_Stink_Bug)_(8212363531).jpg
https://commons.wikimedia.org/wiki/File:Pyrrhocornis_Nymph_-_Order,_Hemiptera._Suborder,_Heteroptera_Family,_probably,_Coreidae_-_Flickr_-_gailhampshire.jpg
https://bugguide.net/node/view/990000/bgimage
https://bugguide.net/node/view/431630/bgimage
https://bugguide.net/node/view/904201
https://www.mindenpictures.com/stock-photo-stink-bug-pentatomidae-nymph-portrait-a-true-bug-of-the-heteroptera-naturephotography-image00762545.html
https://stock.adobe.com/br/images/bug-creontiades-pallidus-hemiptera-miridae-nymph-isolated-on-a-black-background/247036616?as_campaign=ftmigration2&as_channel=dpcft&as_campclass=brand&as_source=ft_web&as_camptype=acquisition&as_audience=users&as_content=closure_asset-detail-page
https://fineartamerica.com/featured/hemiptera-nymph-on-rock-4-douglas-barnett.html
https://fineartamerica.com/featured/hemiptera-nymph-ventral-view-douglas-barnett.html
https://www.mindenpictures.com/stock-photo-stink-bug-pentatomidae-nymph-portrait-a-true-bug-of-the-heteroptera-naturephotography-image00762546.html
https://www.mindenpictures.com/gallery/preview/1058/1472/861/stink-bug-pentatomidae-nymph-portrait-a-true-bug-of-the-heteroptera/0_00762547.html
https://www.gettyimages.pt/detail/fotografia-de-not%C3%ADcias/colorful-stink-bug-nymph-on-lichens-fotografia-de-not%C3%ADcias/939608822
https://www.gettyimages.pt/detail/fotografia-de-not%C3%ADcias/squash-bug-juvenile-plant-feeding-bugs-of-this-fotografia-de-not%C3%ADcias/157901713?adppopup=true
https://www.gettyimages.pt/detail/foto/mottled-shieldbug-or-stink-bug-rhaphigaster-imagem-royalty-free/501885361
https://www.gettyimages.pt/detail/foto/mottled-shieldbug-or-stink-bug-rhaphigaster-imagem-royalty-free/501885361
https://www.gettyimages.pt/detail/foto/carpocoris-purpureipennis-larva-in-the-last-imagem-royalty-free/1032911294?adppopup=true
https://www.gettyimages.pt/detail/foto/western-conifer-seed-bug-larva-on-a-leaf-tip-baden-imagem-royalty-free/1032911316?adppopup=true
https://www.gettyimages.pt/detail/foto/green-shield-bug-larva-hesse-germany-imagem-royalty-free/1284278243?adppopup=true

https://elp.tamu.edu/ipm/bugs/family-pentatomidae-stink-bugs/hemiptera-heteroptera-pentatomidae-podisus-maculiventris-spined-soldier-bug-nymph-a/
https://bugguide.net/node/view/974744/bgimage
https://entomologytoday.org/2017/02/28/invasive-insect-species-a-round-up-of-tools-and-resources/
https://www.wallpaperflare.com/mustha-spinosula-stink-bug-insect-hemiptera-nymph-orange-wallpaper-gtlrx/download/4096x2304
https://www.dpughphoto.com/true_bugs.htm
https://www.dpughphoto.com/green%20stink%20bug%20nymph%20durham%20111421_small.jpg
https://www.dpughphoto.com/green%20stink%20bug%20nymph%20durham%20111421_small.jpg
https://www.dpughphoto.com/true_bugs.htm
https://www.jungledragon.com/image/72448/pyrrhocoris_apterus_-_last_stadium_nymph.html
https://pixels.com/featured/hemiptera-nymph-on-rock-2-douglas-barnett.html
https://commons.wikimedia.org/wiki/File:Reduviidae_-_Sphedanolestes_cingulatus_-_Nymph.JPG
https://hort.extension.wisc.edu/articles/assassin-bugs-and-ambush-bugs-reduviidae/
https://greatestfarms.com/awesome-beneficial-insects-assassin-bug-and-lady-bug-what-you-need/
Instagram: @profpadulla
https://commons.wikimedia.org/wiki/File:Coreidae_nymph.jpg
https://bugguide.net/node/view/1023024
https://www.thoughtco.com/leaf-footed-bugs-family-coreidae-1968621
https://bugguide.net/node/view/436820/bgimage
https://www.jungledragon.com/image/47050/coreidae_nymph_unk._kediri.html
https://www.alamy.es/hoja-footed-bug-familia-coreidae-ninfa-image226199913.html
https://bugguide.net/node/view/903184
https://commons.wikimedia.org/wiki/File:Gonocerus_acuteangulatus_(Coreidae)_-__(nymph),_Elst_(Gld),_the_Netherlands.jpg
https://www.treknature.com/gallery/photo180161.htm
https://bugguide.net/node/view/904201/bgimage
https://bugguide.net/node/view/904202/bgimage
https://www.istockphoto.com/br/foto/bug-do-escudo-verde-ninfa-gm859532680-142020809
https://www.flickr.com/photos/maxwel_rocha/14068835464/galleries/
https://pixabay.com/pt/photos/ninfa-percevejo-inseto-art%C3%B3pode-2574679/
https://www.insetologia.com.br/2014/12/ninfa-de-percevejo-spartocera-em-sao.html
https://www.insetologia.com.br/2018/03/ninfa-de-percevejo-pachylis-em-sao-paulo.html
https://www.insetologia.com.br/2013/04/ninfa-de-percevejo-coreideo-em-sao-paulo.html
https://www.ecoregistros.org/site_br/imagen.php?id=229922
https://www.insetologia.com.br/2013/08/ninfa-de-coreideo-em-goias.html
https://br.pinterest.com/pin/136656169927221644/
https://www.whatsthatbug.com/2006/09/25/corecoris-fuscus-nymph/
https://bugguide.net/node/view/43427
https://www.ecoregistros.org/ficha/Spartocera-sp.
https://foro.infojardin.com/threads/que-chinche-es-esta.15984/
http://veterinaria-animales-mendoza.blogspot.com/2012/03/
https://www.flickr.com/photos/andreaskay/13950734457
https://www.jungledragon.com/wildlife/photos/animalia/arthropoda/insecta/hemiptera/coreidae
http://www.folhapatoense.com/2021/04/27/apos-surgir-em-residencias-inseto-semelhante-ao-barbeiro-e-identificado-pela-vigilancia-ambiental-de-patos/
https://www.whatsthatbug.com/2009/01/30/giant-sweet-potato-bug-nymph/
https://www.agrolink.com.br/problemas/percevejo_1436.html
https://www.whatsthatbug.com/2021/07/06/small-milkweed-bug-nymphs-2/
https://www.whatsthatbug.com/2017/06/24/spined-soldier-bug-nymph/
https://www.whatsthatbug.com/2012/06/20/southern-green-stink-bug-nymph-3/
https://www.whatsthatbug.com/2016/02/29/large-milkweed-bug-nymphs-6/

https://www.whatsthatbug.com/2014/07/18/harlequin-stink-bug-nymphs/
https://www.whatsthatbug.com/2014/05/04/stink-bug-nymphs-everywhere/
https://www.whatsthatbug.com/2011/10/12/harlequin-bug-nymph/
https://www.whatsthatbug.com/2009/01/09/unknown-predatory-red-hemipteran-nymphs-feeding-on-unknown-saturniid-caterpillar-in-south-africa/
https://www.whatsthatbug.com/2007/11/25/large-milkweed-bugs-adults-and-nymphs/
https://bugguide.net/node/view/55555
https://www.whatsthatbug.com/2005/04/04/boxelder-bugs-young-and-nearly-mature-nymphs/
https://guiadoagricultormandriao.blogspot.com/2017/11/encontrei-percevejos-na-minha-horta-o.html
https://bugguide.net/node/view/183996/bgpage
https://bugguide.net/node/view/181270
https://www.scielo.br/j/ne/a/CKMjmRPSz84bJVpd4JcJW9H/?lang=pt&format=pdf
https://australian.museum/learn/animals/insects/bronze-orange-bug/
https://i-heart-bugs.livejournal.com/32790.html
http://magickcanoe.com/blog/?p=508
https://www.flickr.com/photos/e-naturalist/46117728935
https://commons.wikimedia.org/wiki/File:Stink_Bug_Nymph_(7462625830).jpg
https://commons.wikimedia.org/wiki/File:Stink_Bug_nymph_-_Podisus_species,_Leesylvania_State_Park,_Woodbridge,_Virginia.jpg
https://bugguide.net/node/view/1114877
https://bugguide.net/node/view/4253
https://bugguide.net/node/view/540107
https://bugguide.net/node/view/10404
https://br.pinterest.com/pin/231583605826357690/
https://www.mindenpictures.com/stock-photo-stainer-dysdercus-sp-adult-a-true-bug-of-the-heteroptera-suborder-naturephotography-image00762567.html
https://www.dreamstime.com/adult-white-crossed-seed-bug-genus-neacoryphus-image229868739
https://www.insetologia.com.br/p/blog-page_8.html
https://pt.wikipedia.org/wiki/Heteroptera
https://pt.wikipedia.org/wiki/Heteroptera
https://www.insetologia.com.br/p/blog-page_8.html
https://www.flickr.com/photos/wcampelo/6307355796
https://pt.wikipedia.org/wiki/Lygaeoidea
https://www.insetologia.com.br/2017/11/percevejos-coreideos-do-rio-grande-do.html
https://www.insetologia.com.br/2017/09/percevejo-pachylis-na-bahia.html
https://www.amentsoc.org/insects/fact-files/orders/hemiptera-heteroptera.html
https://bugguide.net/node/view/1311512
https://www.biodiversity4all.org/taxa/61267-Heteroptera
https://www.pinterest.es/pin/761530618214219203/
https://sinobug.tumblr.com/post/73581529247/and-a-few-more-true-bugs-of-china-members
https://www.flickr.com/photos/andreaskay/7750538218
https://www.thoughtco.com/true-bugs-order-hemiptera-1968634
https://genent.cals.ncsu.edu/insect-identification/order-hemiptera-suborder-heteroptera/
https://www.easttennesseewildflowers.com/gallery3/index.php/Beetles_Bugs/Copy_of_Milkweed_bug_close
https://sites.psu.edu/sagemckeandpassion/2017/04/07/hemiptera/
https://bugoftheweek.com/blog/2018/10/25/bugs-in-orange-and-black-a-spooky-halloween-trick-or-treat-for-predators-small-and-large-milkweed-bugs-lygaeus-kalmii-and-oncopeltus-fasciatus
https://www.flickr.com/photos/andreaskay/7716153150
https://www.thoughtco.com/red-and-black-bugs-4138391
https://www.thoughtco.com/red-and-black-bugs-4138392
https://www.thoughtco.com/red-and-black-bugs-4138393
https://www.thoughtco.com/red-and-black-bugs-4138394
https://www.thoughtco.com/red-and-black-bugs-4138395
https://www.thoughtco.com/red-and-black-bugs-4138396
https://australian.museum/learn/animals/insects/crusader-bug/

https://www.si.edu/spotlight/buginfo/true-bugs
https://bestlifeonline.com/kissing-bug-news/
https://bestlifeonline.com/kissing-bug-news/
https://prairieecologist.com/2013/02/18/thank-goodness-for-boxelder-bugs/
https://prairieecologist.com/2013/02/18/thank-goodness-for-boxelder-bugs/
https://phys.org/news/2019-04-milkweed-bug-orange-wings-dna.html
https://www.freepik.com/premium-photo/adult-leaf-footed-bug-family-coreidae_19185728.htm
https://www.biodiversity4all.org/taxa/51777-Coreidae
https://www.biodiversity4all.org/taxa/51777-Coreidae
https://www.biodiversity4all.org/photos/19037874
https://www.britishbugs.org.uk/heteroptera/Coreidae/gonocerus_acuteangulatus.html
https://www.flickr.com/photos/andreaskay/7884066184
http://www.cvi.ufscar.br/Imagens/hemiptera/REDUVIIDAE19_gde.jpg
http://www.cvi.ufscar.br/ordens/hemiptera.html
http://www.cvi.ufscar.br/ordens/hemiptera.html
https://www.asturnatura.com/familia/coreidae.html
https://www.asturnatura.com/familia/coreidae.html
https://www.instagram.com/p/Cah0eJEqM3c/
https://www.instagram.com/p/CZ4WOTELmoZ/
https://www.instagram.com/p/BwklfdphbDE/
https://www.instagram.com/p/CPDx-VbsZOJ/
https://www.instagram.com/p/CCQ74fWgEXO/
https://www.instagram.com/p/Cb05FyGpEZH/
https://www.instagram.com/p/CbzwgmIOfMv/
https://www.instagram.com/p/CbzwFWEO3-G/
https://www.instagram.com/p/CbxrCIZLdTV/
https://www.instagram.com/p/CbxqpVTLofr/
https://www.instagram.com/p/CbxqAVprWO9/
https://www.instagram.com/p/CbXPWDHL927/
https://www.instagram.com/p/CbmdBASoRdO/
https://www.instagram.com/p/CbjAWNcLPzO/
https://www.instagram.com/p/CbR6ZDdpTCJ/
https://www.instagram.com/p/CbP7iYpvB7J/
https://www.instagram.com/p/CbPBOtQFq7L/
https://www.instagram.com/p/CbOfNWsAaH1/
https://www.instagram.com/p/CbN7R1csooB/
https://www.instagram.com/p/CbN7R1csooB/
https://www.instagram.com/p/CbN7R1csooB/
https://www.instagram.com/p/CbL4Erjsv_K/
https://www.instagram.com/p/Cbl_Sq6AVyD/
https://www.instagram.com/p/CblJ7LygwjA/
https://www.instagram.com/p/CbFyf5fqPXk/
https://www.instagram.com/p/CbFyf5fqPXk/
https://www.instagram.com/p/CbAOPrAO5kH/
https://www.instagram.com/p/Ca_Q77OLcU-/
https://www.instagram.com/p/Ca7blb7q4n4/
https://www.instagram.com/p/Ca7blb7q4n4/
https://www.instagram.com/p/Ca5nsUrvwS8/

CONSIDERAÇÕES FINAIS

Ao longo desta tese, exploramos profundamente a aplicação de técnicas de aprendizado de máquina na identificação de triatomíneos, vetores de *T. cruzi*, representando um avanço significativo na abordagem tradicional de taxonomia. Com base nos três capítulos apresentados, é possível traçar um panorama abrangente sobre a eficácia e o potencial dessas técnicas para melhorar a vigilância e o controle desses insetos.

No capítulo 1, a aplicação de algoritmos de aprendizado de máquina na identificação automática de vetores da doença de Chagas demonstrou resultados promissores. A capacidade desses algoritmos em lidar com imagens de baixa resolução capturadas por câmeras de celulares comuns ressalta sua relevância para fortalecer os sistemas de vigilância entomológica, especialmente em contextos de declínio na expertise taxonômica. A escolha da rede neural convolucional AlexNet como o algoritmo de melhor desempenho reforça a eficácia das redes neurais pré-treinadas na identificação de vetores de doenças e outros insetos. No entanto, a complexidade da identificação aumenta com o número de classes, exigindo a definição criteriosa de conjuntos de espécies para treinamento e teste dos algoritmos. A modelagem (GLMM) adotada neste estudo contribuiu para uma avaliação mais rigorosa do desempenho dos sistemas de identificação, levando em consideração a dependência entre as observações.

No capítulo 2, avançamos na identificação automática de uma grande diversidade triatomíneos, a partir da construção de um banco de imagens bem representativo das espécies. A aplicação da rede neural convolucional AlexNet, combinada com um banco de imagens abrangente e diversificado, revelou-se promissora ao alcançar altos níveis de acurácia na identificação de uma variedade de espécies, incluindo aquelas de maior importância para a saúde pública. A capacidade da AlexNet em discernir entre espécies como *P. megistus*, *R. prolixus* e *T. brasiliensis*, entre outras, tem o potencial de fortalecer substancialmente as estratégias de vigilância e controle da doença de Chagas. Embora tenhamos obtido sucesso em nossos esforços, a ampliação contínua do banco de dados é crucial para a validade e abrangência deste sistema automatizado de identificação. A inclusão de espécies de diferentes regiões das Américas e a promoção do engajamento comunitário por meio de um aplicativo de acesso aberto são etapas fundamentais para consolidar nossa abordagem inovadora.

Finalmente, no capítulo 3, exploramos o desempenho da rede neural convolucional AlexNet para identificação de ninfas dos vetores da doença de Chagas. Nossos resultados são uma manifestação clara da promissora aplicação de técnicas de aprendizado de máquina para solucionar um desafio crucial na vigilância e controle desses vetores. A acurácia excepcionalmente alta alcançada pela AlexNet, destaca a sua capacidade de distinguir entre as complexas morfologias das ninfas de triatomíneos e outros insetos da subordem Heteroptera, além da discriminação taxonômica entre os gêneros de triatomíneos. A abordagem que empregamos neste estudo apresenta avanços significativos em relação a pesquisas anteriores, uma vez que expande a análise de aprendizado de máquina para o grupo das ninfas de triatomíneos, um aspecto pouco explorado até o momento. A maioria dos estudos anteriores concentrou-se na identificação de espécimes adultos, negligenciando as formas imaturas. Nossos resultados inovadores estabelecem as bases para uma ferramenta capaz de diferenciar ninfas de triatomíneos com precisão, levando em consideração tanto a complexidade taxonômica quanto a importância epidemiológica desses insetos.

Enquanto enfrentamos os desafios na vigilância de vetores da doença de Chagas, nossa pesquisa estabelece um caminho promissor para um sistema de vigilância mais robusto e eficaz. A integração da tecnologia de aprendizado de máquina com a colaboração ativa da comunidade através da ciência cidadã emerge como uma resposta promissora aos obstáculos enfrentados no campo da saúde pública. Acreditamos firmemente que, com a continuidade deste trabalho e a expansão da base de dados, estaremos bem preparados para enfrentar os desafios crescentes na vigilância de vetores da doença de Chagas e contribuir para o controle dessa doença.