Application of Deep Learning to Automated Species Identification Systems

By
Ali Khalighifar

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: A. Townsend Peterson

Rafe Brown

Jorge Soberón

Robert Moyle

Perry Alexander

Date Defended: 12 August 2020

The dissertation committee for Ali Khalighifar certifies that this is the approved version of the following dissertation:

Application of Deep Learning to Automated Species Identification Systems

Chair: A. Townsend Peterson

Date Approved: 17 August 2020

ABSTRACT

Researchers monitor species' populations for a variety of purposes, including surveillance of endangered or rare species for conservation goals, mitigation of public health threats caused by diseases vectored by insects, and threats to native populations by invasive species. Achieving each of these goals is essential to healthy ecosystems and human societies. For instance, species important in biodiversity conservation provide valuable ecosystem services, and disease vectors and invasive species impose heavy economic burdens, along with acute, negative impacts on human health and well-being. As such, the need for robust, efficient, and widely-applicable biodiversity monitoring techniques is at a premium.

Traditional monitoring approaches generally involve labor-intensive data collection processes, with relatively short survey windows and limited spatial coverage. As such, these techniques are often unable to meet the increasing demands of global conservation and public health surveillance. Recent technology advances offer tools, such as audio recording devices, camera traps, and network systems, for transferring digital data, that have revolutionized aspects of biodiversity monitoring. However, while they can provide longitudinal, highly accurate data, these devices generate massive amounts of information that can overwhelm conservation biologists and public health providers.

One solution is to apply real-time, automated systems to process and analyze such data streams. Deep-learning techniques can provide a cyberinfrastructure that can achieve real-time, automated species identifications deriving from such automated devices. They also provide the opportunity to explore, test, and discover unknown or overlooked evolutionary, ecological, and behavioral phenomena regarding target species or regions, making them potentially powerful tools for future basic research in biology. Lastly, as deep-learning techniques can be implemented in

citizen-science platforms, they allow community members to participate in public health and biodiversity science in meaningful and actionable ways.

Here, I present three examples of potential applications of such automated species identification systems. In each example, I explore the potential power of these tools by challenging an advanced deep-learning technique, TensorFlow Inception v3, to identify a set of taxa under different goals, all centered on accurate species identifications. The first chapter of my work delivers a comparison between a deep-learning-, image-based species identification system, and conventional classifiers for insect vectors of Chagas disease in Mexico and Brazil. The second and third chapters focus on the identification of acoustic signals made by taxa, using the trick of converting sounds to spectrograms: images representing auditory features. Specifically, the second chapter demonstrates successful application of a deep-learning model to a diverse clade of closely-related frog species in the Philippines, using single-note mating calls. The third chapter explores the potential for application of automated identification platforms to mosquito species using wingbeat patterns, with emphasis on participation of citizen scientists to improve surveillance of disease vectors.

ACKNOWLEDGMENTS

I would like to give sincere gratitude to my advisor and wonderful friend, Town Peterson, for his continuous support and dedication to my success throughout my PhD. Your excellent mentorship and guidance are the main reasons I am pursuing AI research today. May our ping-pong competitions be close, but may you never truly win.

I would also like to thank my PhD committee members, Rafe Brown, Jorge Soberón, Rob Moyle, and Perry Alexander, as well as former member Craig Martin, for advising me on how to fulfill a multitude of academic goals over the past five years. In particular, I would like to thank Perry for introducing me to experts in computer science and AI, and for financially supporting portions of my PhD through the Blockchain project. Many thanks to Rafe for his significant contributions to my second chapter, for career advice, and for teaching me that frogs are so cool.

Outside of my committee, many colleagues greatly contributed to my success as a PhD student. Without their support, none of this work would have been possible. First, many thanks to Ed Komp for his patience in teaching me how to program like a computer scientist. To Janine Ramsey and Rodrigo Gurgel-Gonçalves, thank you for your entomological insights and taxonomic input to my first chapter. To Johana Goyes Vallejos, a great friend and colleague, many thanks for your expert inputs on frog bioacoustics and behavior to my second chapter. I thank mosquito experts Koffi Ahadji-Dabla, Lindsay Campbell, and Nathan Burkett-Cadena for their taxonomic input and identification of mosquito individuals in my third chapter. I also thank Rich Glor for letting me store mosquitoes in his freezer. Most importantly, many thanks to the friends who patiently helped me collect 350+ mosquitoes: Daniel “Carnalito” Jiménez-García, Town Peterson, Spencer Mattingly, Soheil Shojaei, Fernando Machado-Stredel, Juan Maita-Chamba, Marlon Cobos-Cobos, Claudia Nuñez-Penichet, Daniel Romero-Alvarez, Gaby Valverde-Muñoz, and

Khaleesi and Danerys Salazar. To all of you, I apologize for your mosquito bites and the hot, sticky Kansas nights. To Khaleesi and Danerys in particular, I would never have finished collecting without your blackmail—thanks so much for your encouragement.

I would also like to acknowledge Aagje Ashe, Jaime Keeler, and Rick Evanhoe for being patient with my many questions, and the KU International Student Service for their help in bringing me to Kansas and getting established as an international student in the US.

During my time at KU, I have had the fortune of making many close friends. First, I want to express profound gratitude and love to my best friend, Paige Hansen, for her amazing support while writing my dissertation. To Rosa Salazar-Peterson, thank you so much for making me feel welcome in your house and treating me like a member of your family. To my close friend Spencer Mattingly and my many Iranian friends, especially Amir Barati, Majid Iravani, Soheil Shojaei, and Arash Givchi, thank you for many amazing memories and for always being there during difficult times. To my wonderful friend Mike Wuthrich, thank you for your life and career advice, and for many great conversations over breakfast. I would also like to thank my friends in the KU Ecological Niche Modeling group and KU Biodiversity Institute, especially Daniel Jiménez-García, my officemate Kate Ingenloff, Fernando Machado-Stredel, Javier Torres, Rich Glor, Roberta Marques, Thilina de Silva, Sumudu Fernando, Marlon Cobos-Cobos, Claudia Nuñez-Penichet, Luis Osorio-Olvera, and Luke DeCicco, for helping to create a fun and happy work atmosphere.

Lastly and most importantly, I would like to dedicate this dissertation to my parents and siblings, who have always believed in me, and encouraged me to pursue all my life goals, no matter how big. I miss you all dearly.

CONTENTS

INTRODUCTION

Biodiversity surveillance covers a wide range of goals, including monitoring rare or endangered species for conservation (Barata et al. 2017), discovering new species and understanding their ecological or evolutionary roles in the ecosystem (Brown and Gonzalez 2007, Carvalho et al. 2010, Brown et al. 2015a), and detecting invasive species or disease vectors to mitigate their harms and costs (Bogich et al. 2008, Smolinski et al. 2017). Traditional monitoring approaches generally rely on sending skilled observers or "parataxonomists" to pre-determined sites for collecting data. These methods are costly, laborious, and mostly short-term in nature, limiting monitoring efforts to small-scale and short-term assessments of regions, generally with the exclusion of remote and hard-to-access sites (Ogden 2018, Wang and Gamon 2019). As such, traditional biodiversity survey methods generally suffer from under-sampling through both space and time, as well as human biases related to knowledge of taxa and access to sites.

To address these limitations, researchers have begun to utilize high-tech devices for monitoring species populations (Brown et al. 2002, Webster and Budney 2017), including high-end remote sensors (Papeş et al. 2010, Wang and Gamon 2019), powerful microphone arrays (Blumstein et al. 2011), and high-resolution camera traps (Rowcliffe et al. 2011, Schmeller et al. 2017). In addition to providing more objective, higher-quality data, these devices are able to record metadata associated with digital objects precisely, and in some cases transfer data to a central facility for storage (Hill et al. 2018). More recent technological advances have allowed researchers to monitor biodiversity using more general equipment, including smartphones (Mukundarajan et al. 2017, Nugent 2018), drones (Zhang et al. 2016), and personal cars (Ascensão et al. 2020), which offers the added advantage of opportunities for involvement of citizen scientists. Although these technologies provide researchers with more data with which to address more research questions

and understand complex biodiversity shifts better, high-end devices generate enormous amounts of data that can easily overwhelm researchers (Villanueva-Rivera and Pijanowski 2012, Klein et al. 2015). That is, too often, data are collected, but never analyzed, and analyzing such enormous quantities of data is highly prone to human error.

In addressing this issue, one effective solution is to develop automated, machine-learning-based, real-time species identification systems that are implemented in mobile phone applications, cloud-based web applications, or high-performance clusters. One example of these models are deep neural networks (DNNs), a newly introduced set of machine-learning algorithms that typically outperform other classifiers (Schmidhuber 2015, Smith et al. 2019). DNNs are trainable computational models that comprise several stack of processing layers, each of which learns more abstract information from input data based on information gained by previous layers (Lundervold and Lundervold 2019). These powerful algorithms can efficiently analyze data and reduce their dimensionality, significantly improving the time needed to analyze data and making more information usable for researchers. Convolutional neural networks (CNNs; Yamashita et al. 2018) are a subset of DNNs--they contain many layers and are specialized for image recognition, which makes them ideal for species identification systems. In addition, CNN-based platforms, such as TensorFlow (Abadi et al. 2016), PyTorch (Ketkar 2017), or MXNet (Chen et al. 2015) can allow researchers to deploy flexible, state-of-the-art, deep-learning application to citizen-science-implemented projects, accelerating data processing to improve our understanding of global biodiversity.

Here, I present an exploration of one advanced, deep-learning-based cyberinfrastructure, TensorFlow Inception v3, as regards three biological questions, to test the performance and utility of such models in the fields of biodiversity conservation and public health. The first chapter of this

dissertation focuses on identifying vectors of Chagas disease in Mexico and Brazil, by challenging the model with two groups of cryptic species complexes, *Triatoma dimidiata* and *Triatoma phyllosoma*. Chagas disease, a serious, chronic disease that causes cardiac morbidity and mortality, is a public health threat across the Americas (Rassi et al. 2010); as a consequence, this chapter represents an important early step in creating tools important in mitigating the negative health and economic impacts associated with this disease.

The second chapter demonstrates the potential for accelerating species discovery rates using deep-learning-based models. Here, we used sound spectrogram images representing simple bioacoustic features of advertisement calls (only single notes) associated with 41 closely-related frog species in the Philippines (Brown et al. 2012, Brown et al. 2015b) to test TensorFlow's ability to discriminate among species, and to test its performance in discovering and "flagging" recordings of novel species that are not present in the reference library. As the Philippines represents a significant biodiversity hotspot, with 110+ amphibian species (90% of which are endemic to this archipelago) and showing one of the highest rates of species discovery in the world (Rowley et al. 2009, Diesmos et al. 2015), this work is an important step to conserving known species diversity and identifying unknown biodiversity.

Finally, the third chapter outlines the availability of infrastructure for detecting invasive species of public health interest, benefiting from citizen science involvement. In this chapter, we used recordings of mosquito wingbeats collected by built-in smartphone microphones to identify mosquito species native to Douglas County, Kansas. A specific challenge of interest was to assess the ability of the identification algorithms to detect two potential novel invasive species of public health importance, *Aedes aegypti* and *Anopheles gambiae*, not currently present in the study area. Because mosquito-borne diseases are responsible for hundreds of thousands of deaths around the

world each year, and because mosquito species distributions are predicted to shift significantly with climate change (Reiter 2001), this pilot study explores the potential utility of citizen-science-based automated identification systems to monitor the spread of harmful diseases to northern latitudes.

CHAPTER 1. Deep learning algorithms improve automated identification of Chagas disease vectors

1.1 Introduction

Vector-borne diseases, such as malaria, Zika, dengue fever, Chagas disease, and chikungunya, impose enormous economic and social burdens on public health (Sinkins and Gould 2006), and rank among the most serious threats to human health (Nauen 2007). An important element in predicting risk of such diseases, and developing strategies to control them, is detailed surveillance of arthropod vector species (Mukundarajan et al. 2017). Identifying vectors can be challenging, requiring significant training and increasingly uncommon taxonomic expertise (Drew 2011), a phenomenon termed the "taxonomic impediment" (de Carvalho et al. 2007). As a result, providing broad regions with necessary arthropod vector identification expertise by traditional means is often not feasible given political boundaries, logistical obstacles, and variability of country-specific human health resources (Schofield et al. 2006). One promising potential solution involves developing automated species identification systems that can aid in meeting these important challenges.

Chagas disease occurs in most countries in the Americas, causing serious cardiac morbidity and mortality among infected individuals over years or decades if untreated (Coura and Viñas 2010). The disease is caused by the unicellular parasitic protozoan, *Trypanosoma cruzi* (Trypanosomatida: Trypanosomatidae), transmitted mainly through feces of blood-sucking true bugs belonging to the hemipteran subfamily Triatominae (Hemiptera: Reduviidae) (Coura and Viñas 2010). Estimates suggest that 8 million people exhibit symptoms of the disease, and that >100 million people are at risk of infection worldwide (Coura and Viñas 2010, WHO 2018). Given failings in many public health systems across the Americas, epidemiological surveillance for

Chagas disease is either minimal (e.g., in Mexico) or incomplete (e.g., in Brazil), and diagnosis and treatment are often too late and therefore less effective (Martins-Melo et al. 2012, Ramsey et al. 2014).

Many efforts have tested diverse inferential approaches in the challenge of identifying insects accurately. Recently, a new set of machine-learning classifiers, called deep neural networks (DNNs), have been designed; they are similar to artificial neural networks (ANNs; Dayhoff and DeLeo 2001), but with multiple hidden layers between the input and output layers, instead of having only three layers total (i.e., input, one hidden layer, and output). The high capability of DNNs has resulted in widespread use of these techniques in other such inference challenges (Schmidhuber 2015). DNNs have outperformed traditional methods (e.g., linear discriminant analyses) in various classification tasks (Schmidhuber 2015). Here, we used TensorFlow (Abadi et al. 2016), an open-source software platform, representing the most recent addition to the deep-learning toolbox (Rampasek and Goldenberg 2016; Google Brain Team; https://research.google.com/teams/brain/), to explore the challenge of automated Chagas disease vector identification.

The purpose of this study was to test whether deep-learning techniques can improve abilities to identify Chagas disease vectors of Mexico and Brazil successfully, using a collection of digital images of triatomines from our previous work (Gurgel-Gonçalves et al. 2017). To address this question, we designed an automated, DNN-based species identification system for 12 Mexican and 38 Brazilian species, which we tested using images withheld from the calibration process. This study thereby lays a foundation for many automated identification capabilities in medical entomology, and more broadly in biodiversity science.

1.2 Materials and Methods

We tested whether state-of-the-art, deep-learning techniques (TensorFlow; Abadi et al. 2016) can successfully identify digital images of 12 Mexican triatomine species and 38 Brazilian triatomine species. We implemented four steps: (1) adapt and configure TensorFlow to the needs of this analysis; (2) create different sets of images, based on different levels of automated pre-processing, from available images, as inputs; (3) apply and test TensorFlow under diverse circumstances; and (4) use secondary information (i.e., distributional information) to reduce numbers of species in our identification exercises. We implemented all analyses in Python (version 2.7, available at http://www.python.org/). These steps are described in greater detail below.

Previously (Gurgel-Gonçalves et al. 2017), we presented work on automated identification of triatomine species from Mexico (12 species) and Brazil (39 species). That work involved full automation of processing of images before analysis into a set of landmarks and distances among them, and then analysis using traditional statistical classifiers, such as linear discriminant analysis (Fisher 1936) and artificial neural networks (Dayhoff and DeLeo 2001). Identifying the Mexican species was a particular challenge to our classifiers, and of course we were interested in improving identification rates for Brazilian species also. Therefore, we decided to use more advanced classifiers for both Mexican and Brazilian triatomines. To be able to compare our results, we used the same species and same images for this study (Appendix 1, Appendix 2); however, given small numbers of images available for *Eratyrus mucronatus* (Stål) (Hemiptera: Reduviidae) (n=11), we removed this species from the Brazilian pool.

1.2.1 Photographs as input

In our previous study, we had observed that most triatomine images then available were of poor quality, such that the resulting identifications were not reliable. Consequently, project personnel designed a photo apparatus that would permit taking consistent, high quality, repeatable photographs of triatomines at low cost (Gurgel-Gonçalves et al. 2017). This apparatus allows users to employ iPods or cell phones for taking photographs. Team members photographed triatomine species from various institutions across Mexico (Centro Regional de Investigación en Salud, Instituto Nacional de Salud Pública México; Laboratorio Estatal de Salud Pública de Guanajuato; Universidad Autónoma Benito Juárez, Oaxaca; Universidad Autónoma de Nuevo León, Monterrey), and Brazil (Universidade de Brasília, FIOCRUZ Bahia, FIOCRUZ Rio de Janeiro, FIOCRUZ Minas Gerais, Universidade Estadual Paulista), using the 5th and 6th generations of iPods. Detailed information about the apparatus and protocols for capturing images is provided in Gurgel-Gonçalves et al. (2017).

The photographs that we used as our "reference library" for our identifications are of high-quality, and are openly available at http://dx.doi.org/10.5061/dryad.br14k; they are described in great detail by Gurgel-Gonçalves et al. (2017). These photographs are consistent in orientation, background color, pixel resolution, and quality, which made automation of their processing feasible. Subsequent steps were automated (Gurgel-Gonçalves et al. 2017), including removing (digitally) the background from the images, and identifying the body edge, yielding two sets of images (raw photographs, and images cleaned to remove background) to understand how deep-learning would perform with different quality levels of images as input (Figure 1). The processing step was automated fully and the program code is openly available in the appendices of our

previous paper (Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.br14k; Gurgel-Gonçalves et al. 2017).

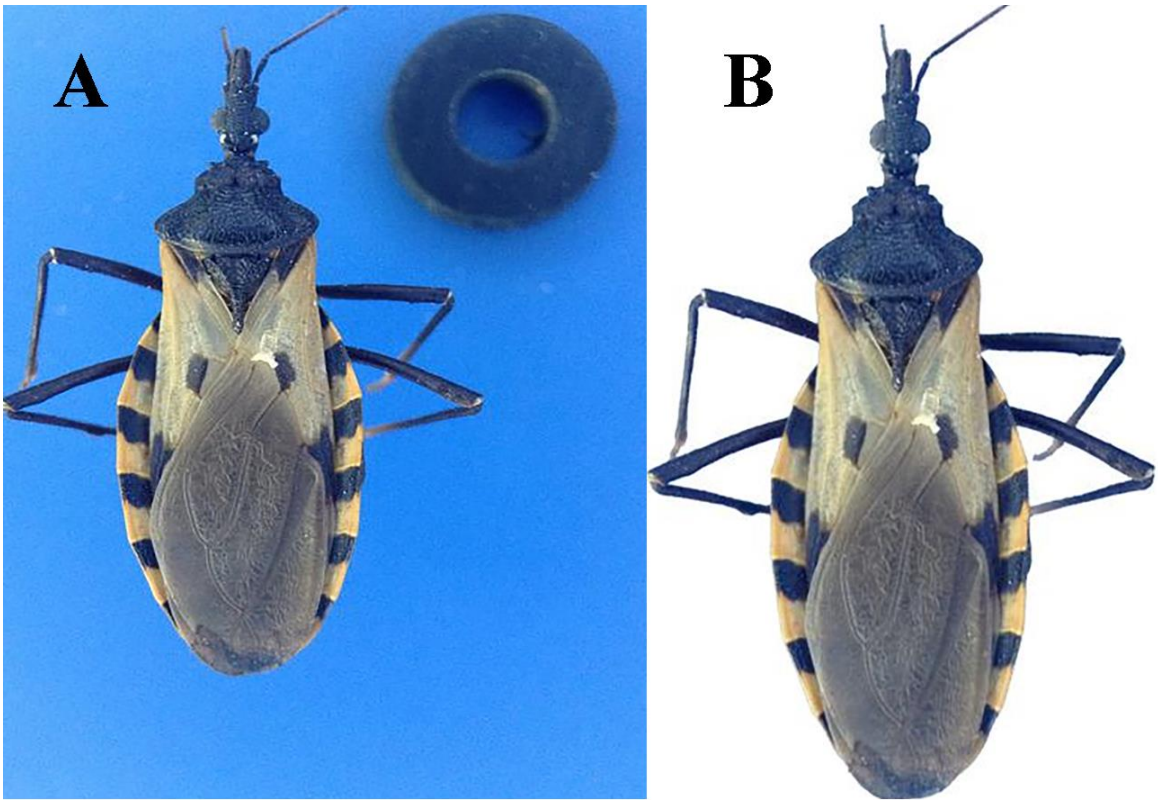

Figure 1. An example image of an individual of *Triatoma dimidiata* Hg1. (A) raw image, (B) final image with background removed digitally.

1.2.2 TensorFlow

TensorFlow is an open-source software platform (Abadi et al. 2016) that supports deep-learning research and applications, having been released by researchers at Google (Google Brain Team; https://research.google.com/teams/brain/). We adapted and configured TensorFlow for application and use in our project via Python. We adjusted two parameters from their default values: validation percentage and number of training steps. Although the default setting for validation percentage was 10%, given small numbers of images available for some species, we

had to increase this number to 16%. For number of training steps, although the default value was 4000 steps, we explored different numbers and compared the results to find the optimum balance between computing time and classification efficiency.

To evaluate the classifier, it is necessary to set some images aside from classifier development entirely. Given the limited numbers of photographs available for some of the species, we used a leave-one-out cross-validation approach for evaluation. That is to say, as a trade-off between maximizing numbers of samples for training the classifier and optimizing computing time, in each round, we left out *n* images, or one image from among the *n* species in that country, with which to evaluate the model.

1.2.3 Identifying faunal subsets based on distributional information

Generally, classifier accuracy decreases as the number of classes that must be distinguished increases. In the preliminary work, triatomine species were immediately separated into two independent groups; those occurring in Mexico and those in Brazil. This separation is natural since almost no triatomine species occur naturally in both countries, exception for *Panstrongylus rufotuberculatus* (Champion) (Hemiptera: Reduviidae). Independent classifiers were created for the two groups. Within these two groups, we applied the same technique on a finer scale, to determine if using a specialized classifier that discriminates only among species known to be present near the location at which a sample is taken will improve the ability to correctly classify triatomine samples. We used the same binary maps as in our previous study (Gurgel-Gonçalves et al. 2017), summarizing potential geographic distributions for each of the 12 Mexican and 38 Brazilian species. Ecological niche modeling techniques and data sources are described in detail in Gurgel-Gonçalves et al. (2012), Ramsey et al. (2015), and Gurgel-Gonçalves et al. (2017).

Hence, in Python 2.7, we created 399 and 2019 local fauna sets for the Mexican and Brazilian triatomine species, respectively, using the distributional information to create subsets of species that co-occur at the same sites. We trained and tested the classifier employing samples restricted to those local faunas. The range of numbers of species in local fauna subsets was 2-11 for Mexican species and 2-12 for Brazilian species.

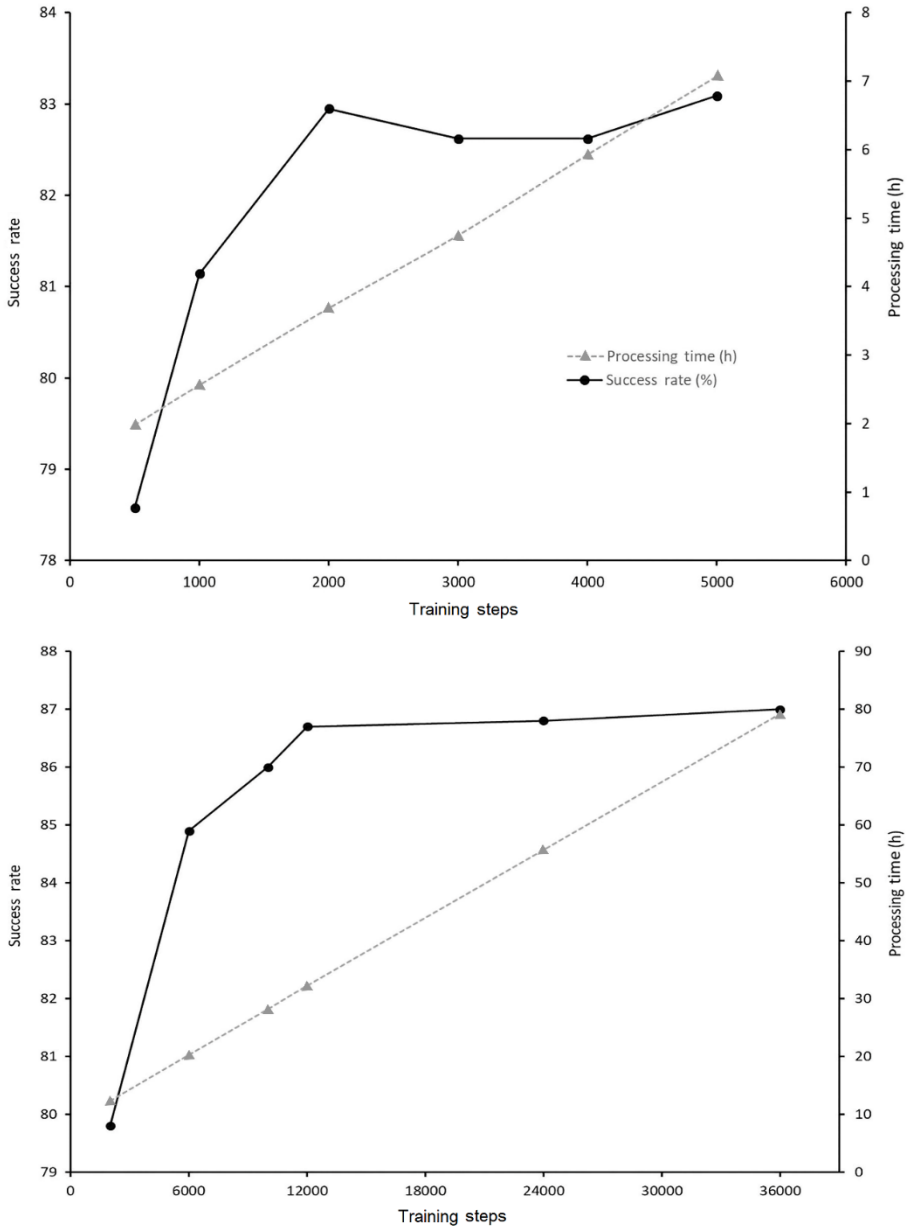

Figure 2. Effects of different numbers of training steps on identification rate (%) and processing time (h) using TensorFlow for Mexican (top) and Brazilian (bottom) triatomine species.

1.3 Results

As a preliminary step, we tested identification success for raw versus cleaned images for Mexican species, to see whether TensorFlow shows reduced correct identification rates using raw images. The identification rate for raw images was almost the same as that obtained using the cleaned images (i.e., 82.9% overall identification rate for raw images, versus 83.0% for cleaned images); since our goal was to compare the results of TensorFlow with those deriving from the statistical classifiers used in our previous study (Gurgel-Gonçalves et al. 2017), for all succeeding analyses in this study, we elected to use cleaned images, only noting the potential for simpler processing in future work.

We calibrated our models for Mexican and Brazilian species using different numbers of training steps, and considered two factors to find optima for training TensorFlow (Figure 2): (1) correct identification rate, which improved up to a certain point, but then reached a plateau, beyond which we saw no significant improvement, and (2) processing time (in hours), which showed a positive linear relationship with number of training steps. For the Mexican dataset, with 12 species, 2000 emerged as an optimum number of training steps, which resulted in an 83.0% overall identification rate (Figure 2); the largest gap between the processing time trace and identification rate occurred following this number of steps. Above this number of training steps, subtle improvements in identification rate were possible (up to 83.1%), but processing time was significantly longer (7 hours instead of 3.5 hours). Based on the same logic, we chose 12,000 training steps as an optimum number for Brazilian species, at which we achieved an 86.7% overall identification rate (Figure 2). The Brazilian dataset had a plateau that was more obvious: overall identification rate barely reached a maximum 87.0%, yet processing time increased from 30 to 80

hours. It should be noted that the amount of time discussed above is only for training TensorFlow over all images in the reference library with multiple replicate analyses; for identification of a single image, it only takes a few seconds.

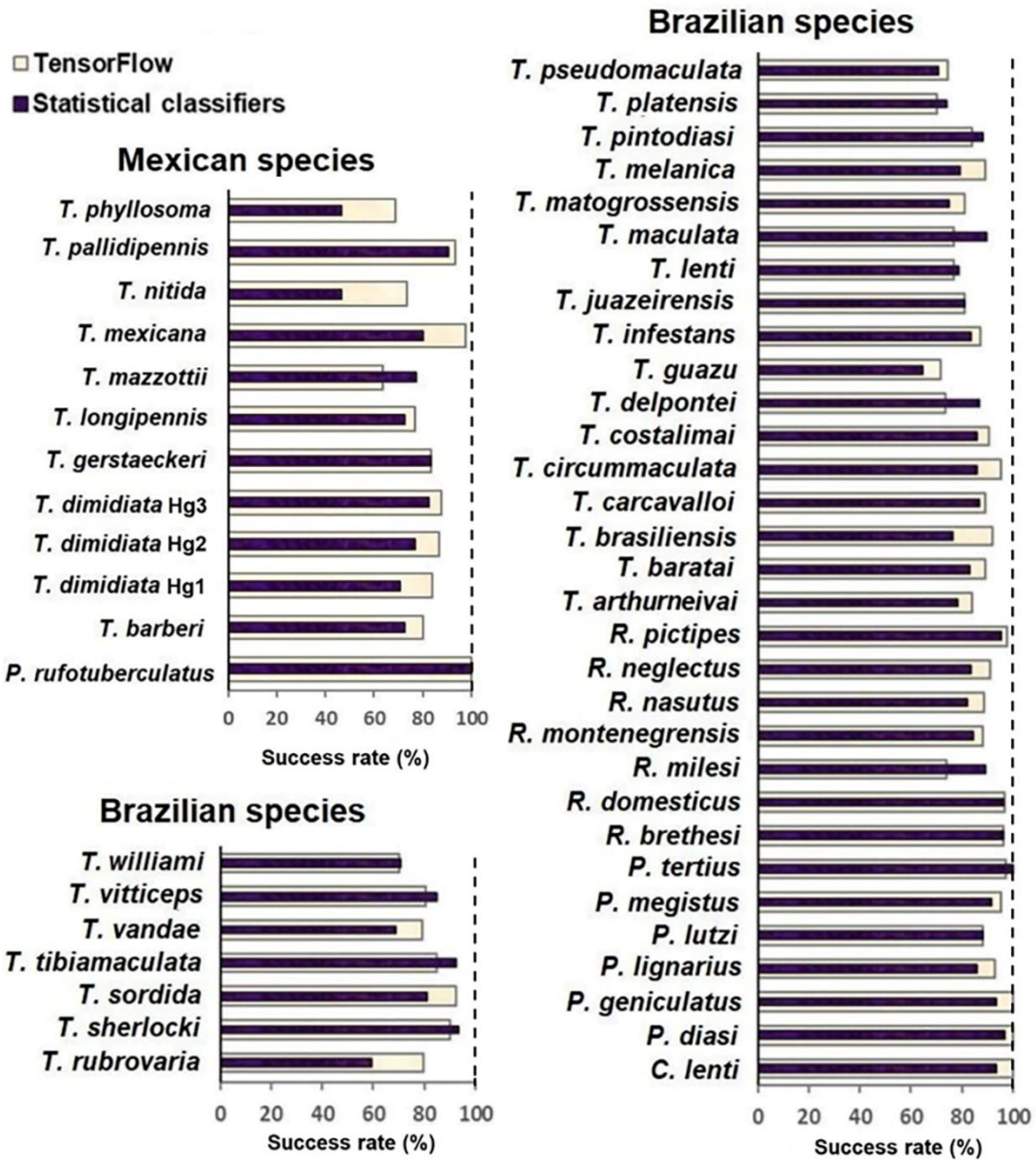

Figure 3. Comparison of identification rates (%) between TensorFlow and statistical classifiers (Gurgel-Gonçalves et al. 2017) at species level for Mexican and Brazilian triatomines.

We plotted and compared the results of TensorFlow with those from our previous study (Gurgel-Gonçalves et al. 2017), at the species level (Figure 3). We applied a paired *t*-test on outcomes, which produced *p*-values of 0.028 and 0.025 for Mexican and Brazilian triatomines, respectively; given that the *p*-values were <0.05, we concluded that the DNN-based results are statistically significantly better than our previous results. For Mexican species, we noted improvement in identification rate for 9 out of 12 species using TensorFlow (Figure 3). Although *T. mazzottii* Usinger correct identification rates declined from 77.3% to 63.6% using deep learning, we noted significant improvement for the *T. dimidiata* (Latreille) complex (*T. dimidiata* Hg1 increased from 70.5% to 84.1%, *T. dimidiata* Hg2 from 76.7% to 86.7%, and *T. dimidiata* Hg3 from 82.5% to 87.5%), *T. mexicana* (Herrich-Schaeffer) from 80.0% to 97.8%, and *T. barberi* from 72.4% to 80.0%. Most notably, we saw improvements for *T. nitida* Usinger and *T. phyllosoma* (Burmeister, 1835), which were very challenging for our previous classifiers, from 46.7% to 73.3% and 46.6% to 69.0%, respectively. For Brazilian species, identification rate improved for 23 out of 38 species, including 6 with major improvements: *T. vandae* Carcavallo, Jurberg, Rocha, Galvão, Noireau & Lent from 69.0% to 79.3%, *T. sordida* (Stål, 1859) from 81.2% to 92.7%, *T. rubrovaria* (Blanchard, 1843) from 59.3% to 79.6%, T. melanica Neiva & Lent from 79.3% to 89.3%, *T. circummaculata* (Stål) from 85.7% to 95.5%, and *T. brasiliensis* Neiva from 76.6% to 92.2%; identification rates for 5 out of 38 species stayed the same (Figure 3). A highlight was that *T. rubrovaria* had shown the weakest classification success with statistical classifiers, but with deep learning improved dramatically. However, 3 species showed major declines: *R. milesi* Carcavallo, Rocha, Galvão & Jurberg (Hemiptera: Reduviidae) from 89.2% to 74.4%, *T. delpontei* Romana & Abalos from 86.7% to 74.2%, and *T. maculata* (Erichson) from 89.7% to 77.5%, compared with the results of the statistical classifiers. The confusion matrices for both Mexican

and Brazilian species, summarizing numbers of images for each species identified correctly or misidentified as another species, are presented in the supplementary data of this paper.

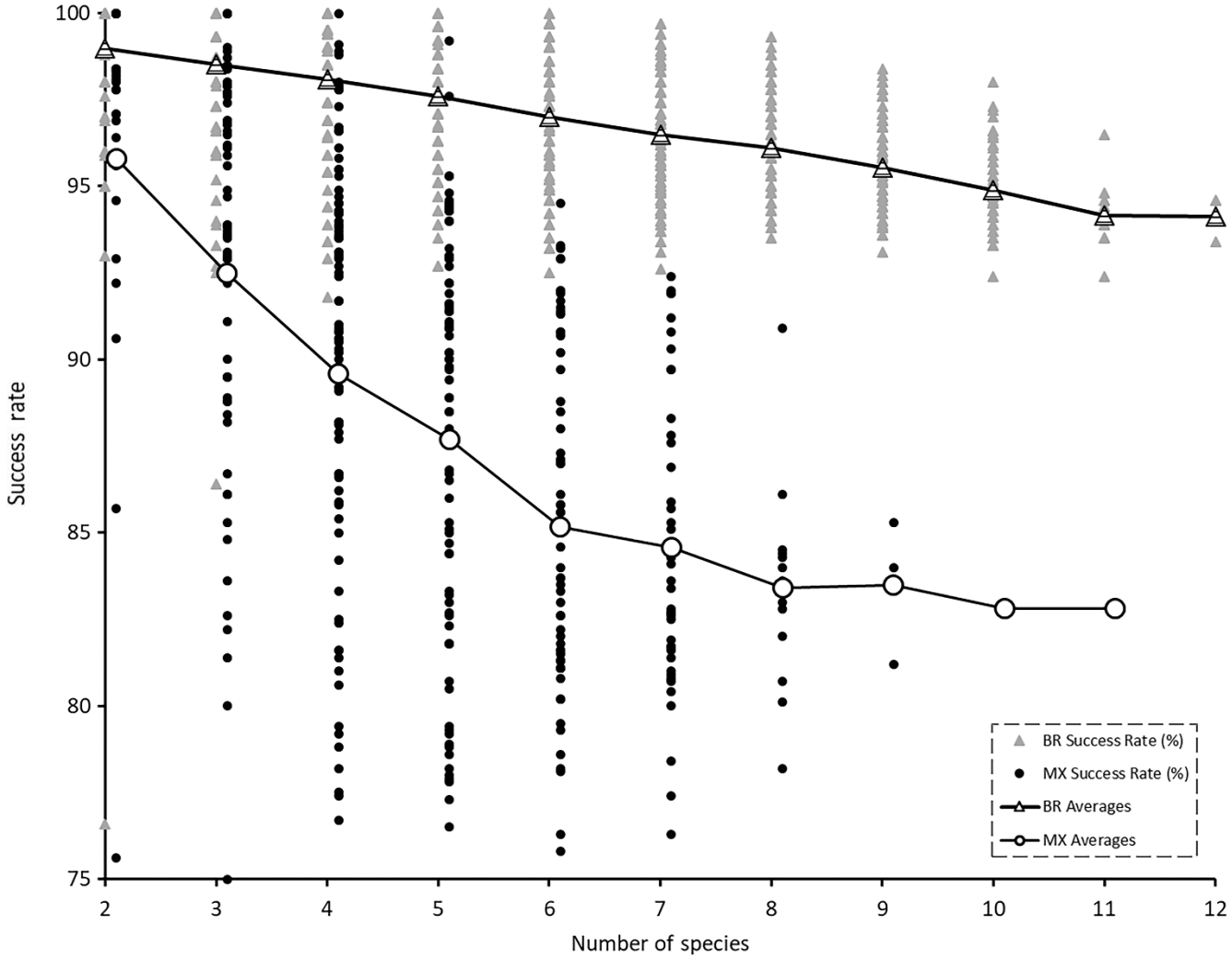

Figure 4. Summary of identification rates (%) for various combinations of species: 2-11 different species combinations for Mexican faunas with 399 faunal subsets, and 2-12 species combinations for Brazilian faunas with 2019 faunal subsets, using the distributional information. Gray triangles represent average identification rates for each Brazilian local fauna, and large black triangles show the overall average of identification rate for each fauna size. Black dots represent average identification rates for each Mexican local fauna, and large blank circles show the overall average of identification rate for each fauna size.

1.3.1 Faunal subsets

Because classifier accuracy improves when fewer classes are compared, to reduce numbers of species being compared, we used geographic information for the distributions of Mexican and Brazilian triatomines to refine identification efforts, testing 399 and 2019 faunal subsets for the two countries, respectively. For Mexican species, faunal subsets ranged 2-11 species (Figure 4). TensorFlow was able to increase the overall identification rate for comparing two species up to 95.8%; average identification rates for most local faunas (25 out of 31) were in the range of 95-100%. As expected, increasing the number of species compared, the overall identification rate declined gradually, and leveled off at around 83.0% for 8 to 11 Mexican species. For Brazilian species, the overall identification rate for comparing two species was 98.9% (Figure 4). Adding more species to the Brazilian pool, the overall identification rate decreased to a plateau of about 94.0% with 11 species. For 1853 out of 2019 Brazilian local fauna subsets, TensorFlow's identification performance was 95% or better.

1.4 Discussion

Chagas disease causes serious cardiac morbidity and mortality among infected individuals over the course of years or decades (Coura and Viñas 2010), yet epidemiological surveillance, diagnosis, and treatment are often late or inefficient owing to shortage of expertise and failings in many public health systems (Martins-Melo et al. 2012, Ramsey et al. 2014). Arriving at such diagnoses depends rather crucially on vector awareness and correct identification, which is the focus of our work. To that end, we have explored deep neural networks (Schmidhuber 2015) to automate several key inferential tasks. We successfully identified both Mexican and Brazilian triatomine species with considerable improvement in overall identification rates, compared to our

previous results (Gurgel-Gonçalves et al. 2017). Incorporating distributional information for species allowed us to reduce local fauna sizes in analyses, making identification easier, which increased overall identification rates for local faunas still more.

A crucial advantage of using TensorFlow was eliminating the need for pre-processing images for our identification system. That is, we compared TensorFlow classification performance based on the cleaned and raw images for the Mexican species, which were the most challenging to our statistical classifiers (Gurgel-Gonçalves et al. 2017). TensorFlow achieved almost the same overall identification rate using the raw images, which opens important future possibilities, such as developing identification based on photos taken by local residents with mobile phone cameras. Including geographic information to create local faunas or faunal subsets (Gurgel-Gonçalves et al. 2012, Ramsey et al. 2015) significantly improved our results, reaching an average accuracy of 95.8% for Mexican triatomines and 98.9% for Brazilian triatomines. Another advantage of TensorFlow is that it is open source software: TensorFlow's flexible architecture allows easy adaptation and deployment of the software on different platforms including desktops, clusters of servers, edge devices, and mobile phones (https://www.tensorflow.org/).

1.4.1 Deep neural networks

Applying TensorFlow to our dataset presented some challenges, a major one was the number of images available for each species. To the best of our knowledge, this is the first study using TensorFlow to identify triatomines. To take advantage of deep learning algorithms to improve identification rates at the species level, we would ideally have access to many more images for each species. We are working along several lines to build such photographic resources, but clearly small sample sizes will be a continuing challenge in this work. A second major

challenge, a consequence of the first, was the model validation method. We would have been happier with our results if we had enough images for each species to follow *k*-fold cross validation approaches (Kohavi 1995), instead of leave-one-out approaches to evaluate our model, as the former offer greater independence between calibration data and evaluation datasets.

Photo quality could constitute a limitation in future analysis. These problems could derive from characteristics of the photographed specimens, such as how dirty it is, the mounting orientation of the specimen, or whether or not the specimen has a pin through its thorax. Other variables that must be considered before full implementation, such as in a "citizen scientist" setting, with photographs captured with ordinary mobile phone cameras, could be evenness of lighting, distance between camera and insect, and how well focused is the image. Future experiments will evaluate the degree to which these factors compromise identification ability, to verify how robust and reliable our automated identifications can be under different circumstances.

1.4.2 Problem species

Although we were impressed by the improved correct identification rates that we obtained, several "problem species" remained, which need to be discussed in more detail. To identify "problem species", we considered a threshold for each dataset relative to overall country-specific identification rates. For Mexican species, we chose 70.0% as the threshold, which resulted in having two problem species, *T. mazzottii* and *T. phyllosoma*, with 63.6% and 69.0% overall identification rates, respectively. Total sample size for *T. mazzottii* was 22 images, of which 8 were misidentified, and for *T. phyllosoma* of 58 images, 18 were misidentified. Out of 8 incorrectly identified samples for *T. mazzottii*, 2 were identified as *T. longipennis* Usinger, and 6 as *T. phyllosoma*. For the 18 misidentified records of *T. phyllosoma*, 2 were identified as *T. pallidipennis*

Stål, 8 as *T. mazzottii*, and 8 as *T. longipennis*. The difficulty of TensorFlow to identify these species may be related to the morphological similarity between them, as they all belong to the *phyllosoma* complex (Lent and Wygodzinsky 1979). However, neither morphological identity nor more recent evolution of the *dimidiata* complex haplogroups affected the ability to distinguish among them. Future experiments with more images should improve correct identification rates.

For Brazilian triatomine species, we selected a 75.0% correct identification threshold, below which five species had lower identification accuracy of 70-74%: *R. milesi*, *T. delpontei*, *T. guazu* Lent & Wygodzinsky, *T. platensis* Neiva, and *T. williami* Galvão, Souza & Lima. All 7 incorrectly identified images of *T. delpontei* were identified as *T. platensis*; other problem species were confused with 2-4 other species. Again, TensorFlow limitation to distinguish between *T. delpontei* and *T. platensis* may be related to morphological similarities (Monteiro et al. 2018).

1.4.3 Future perspectives

In the future, we anticipate deployment of these systems more broadly in terms of both user communities and species under study. The ability to use raw images makes the process of collecting images from different sources (e.g., public health service personnel, citizen scientists) feasible, while maintaining the high accuracy of identification. Our next steps will make this technology available for photographs captured with ordinary mobile phone cameras, and thus broadly accessible to all via a mobile-phone application. The system will (1) permit distinguishing triatomines from non-triatomines, and (2) extend triatomine taxonomic coverage to the whole of the Americas. Extension of such approaches to other medically important arthropod groups is eminently feasible, and could be implemented with relatively little effort, at least in terms of computation. Taxa such as ticks, in particular, are significant vectors of many human and livestock

pathogens (Parola et al. 2005), and could be the most suitable candidates for immediate future studies as they are larger, relatively two-dimensional and can be characterized via a single dorsal-view image. Extending the methodology to mosquitoes and other dipterans may not be too far off (Giordani et al. 2017).

1.5 Acknowledgments

We thank the Office of the Provost, of the University of Kansas, and the Consejo Nacional de Ciencia y Tecnologia (Mexico; CONACyT #261006 to JMRW) for their support of this project. We also thank Rick Evanhoe and Riley Epperson for technical support.

1.6 References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Coura, J. R., and P. A. Viñas. 2010. Chagas disease: A new worldwide challenge. Nature 465: S6-S7.

Dayhoff, J. E., and J. M. DeLeo. 2001. Artificial neural networks. Cancer 91: 1615-1635.

de Carvalho, M. R., F. A. Bockmann, D. S. Amorim, C. R. F. Brandão, M. de Vivo, J. L. de Figueiredo, H. A. Britski, M. C. de Pinna, N. A. Menezes, and F. P. Marques. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evol. Biol. 34: 140-143.

Drew, L. W. 2011. Are we losing the science of taxonomy? As need grows, numbers and training are failing to keep up. BioScience 61: 942-946.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. Hum. Genet. 7: 179-188.

Giordani, B., A. Andrade, E. Galati, and R. Gurgel-Gonçalves. 2017. The role of wing geometric morphometrics in the identification of sandflies within the subgenus *Lutzomyia*. Med. Vet. Entomol. 31: 373-380.

Gurgel-Gonçalves, R., C. Galvão, J. Costa, and A. T. Peterson. 2012. Geographic distribution of Chagas disease vectors in Brazil based on ecological niche modeling. J. Trop. Med. 2012: 705326.

Gurgel-Gonçalves, R., E. Komp, L. P. Campbell, A. Khalighifar, J. Mellenbruch, V. J. Mendonça, H. L. Owens, K. de la Cruz Felix, A. T. Peterson, and J. M. Ramsey. 2017. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: The Virtual Vector Lab. PeerJ 5: e3040.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, pp. 1137-1145, IJCAI, Montreal, Canada.

Lent, H., and P. Wygodzinsky. 1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bull. Am. Mus. Nat. Hist. 163: 123-520.

Martins-Melo, F. R., C. H. Alencar, A. N. Ramos, and J. Heukelbach. 2012. Epidemiology of mortality related to Chagas disease in Brazil, 1999–2007. PLoS Negl. Trop. Dis. 6: e1508.

Monteiro, F. A., C. Weirauch, M. Felix, C. Lazoski, and F. Abad-Franch. 2018. Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. Adv. Parasitol. 99: 265-344.

Mukundarajan, H., F. J. H. Hol, E. A. Castillo, C. Newby, and M. Prakash. 2017. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. eLife 6: e27854.

Nauen, R. 2007. Insecticide resistance in disease vectors of public health importance. Pest Manag. Sci. 63: 628-633.

Parola, P., C. D. Paddock, and D. Raoult. 2005. Tick-borne rickettsioses around the world: Emerging diseases challenging old concepts. Clin. Microbiol. Rev. 18: 719-756.

Rampasek, L., and A. Goldenberg. 2016. TensorFlow: Biology's gateway to deep learning? Cell Syst. 2: 12-14.

Ramsey, J. M., M. Elizondo-Cano, G. Sanchez-González, A. Peña-Nieves, and A. Figueroa-Lara. 2014. Opportunity cost for early treatment of Chagas disease in Mexico. PLoS Negl. Trop. Dis. 8: e2776.

Ramsey, J. M., A. T. Peterson, O. Carmona-Castro, D. A. Moo-Llanes, Y. Nakazawa, M. Butrick, E. Tun-Ku, K. d. la Cruz-Félix, and C. N. Ibarra-Cerdeña. 2015. Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. Mem. Inst. Oswaldo Cruz 110: 339-352.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. Neural Netw. 61: 85-117.

Schofield, C. J., J. Jannin, and R. Salvatella. 2006. The future of Chagas disease control. Trends Parasitol. 22: 583-588.

Sinkins, S. P., and F. Gould. 2006. Gene drive systems for insect disease vectors. Nat. Rev. Genet. 7: 427-435.

1.7 Appendix

Appendix 1. Summary of species analyzed, sample size in terms of numbers of photographs, and successful identification rates (%), in our previous analyses (Gurgel-Gonçalves et al. 2017) versus using deep learning techniques, for the 12 Mexican species in this study. Hg1 – Hg3 represent three distinct haplogroups of *Triatoma dimidiata* complex.

Species	**Sample size**	**Statistical classifiers success rate (%)**	**Deep learning success rate (%)**
Panstrongylus rufotuberculatus (Champion, 1899)	13	100.0	100.0
Triatoma barberi Usinger, 1939	30	72.4	80.0
Triatoma dimidiata (Latreille, 1811) Hg1	44	70.5	84.1
Triatoma dimidiata (Latreille, 1811) Hg2	30	76.7	86.7
Triatoma dimidiata (Latreille, 1811) Hg3	40	82.5	87.5
Triatoma gerstaeckeri (Stål, 1859)	12	83.3	83.3
Triatoma longipennis Usinger, 1939	52	72.5	76.9
Triatoma mazzottii Usinger, 1941	22	77.3	63.6
Triatoma mexicana (Herrich-Schaeffer, 1848)	45	80.0	97.8
Triatoma nitida Usinger, 1939	15	46.7	73.3
Triatoma pallidipennis Stål, 1872	44	90.7	93.2
Triatoma phyllosoma (Burmeister, 1835)	58	46.6	69.0
Total sample size and average success rates	405	80.3	83.0

Appendix 2. Summary of species analyzed, sample size in terms of numbers of photographs, and successful identification rates (%), in our previous analyses (Gurgel-Gonçalves et al. 2017) versus using deep learning techniques, for the 38 Brazilian species assessed in this study.

Species	Sample size	Statistical classifiers success rate (%)	Deep learning success rate (%)
Cavernicola lenti Barrett & Arias, 1985	32	93.3	100.0
Panstrongylus diasi Pinto & Lent, 1946	30	96.7	100.0
Panstrongylus geniculatus (Latreille, 1811)	45	93.3	100.0
Panstrongylus lignarius (Walker, 1873)	30	85.7	93.3
Panstrongylus lutzi Neiva & Pinto, 1923	35	88.2	88.2
Panstrongylus megistus Burmeister, 1835	85	91.7	96.5
Psammolestes tertius Lent & Jurberg, 1965	36	100.0	97.2
Rhodnius brethesi Matta, 1919	28	96.4	96.4
Rhodnius domesticus Neiva & Pinto, 1923	30	96.3	96.7
Rhodnius milesi Carcavallo, Rocha, Galvão & Jurberg, 2001	43	89.2	74.4
Rhodnius montenegrensis Rosa et al. 2012	43	84.6	88.1
Rhodnius nasutus Stål, 1859	80	82.2	88.8
Rhodnius neglectus Lent, 1954	67	83.3	90.9
Rhodnius pictipes Stål, 1872	43	95.3	97.7
Triatoma arthurneivai Lent & Martins, 1940	32	78.1	84.4
Triatoma baratai Carcavallo & Jurberg, 2000	29	82.8	89.7
Triatoma brasiliensis Neiva, 1911	64	76.6	92.2
Triatoma carcavalloi Jurberg, Rocha & Lent, 1998	38	86.8	89.5
Triatoma circummaculata (Stål, 1859)	22	85.7	95.5
Triatoma costalimai Verano & Galvão, 1958	64	85.7	90.6
Triatoma delpontei Romana & Abalos, 1947	31	86.7	74.2
Triatoma guazu Lent & Wygodzinsky, 1979	29	64.3	72.4
Triatoma infestans (Klug, 1834)	55	83.3	87.3
Triatoma juazeirensis Costa & Felix, 2007	20	81.0	81.0
Triatoma lenti Sherlock & Serafim, 1967	39	78.9	77.5

Species	**Sample size**	**Statistical classifiers success rate (%)**	**Deep learning success rate (%)**
Triatoma maculata (Erichson, 1848)	40	89.7	77.5
Triatoma matogrossensis Leite & Barbosa, 1953	33	75.0	81.8
Triatoma melanica Neiva & Lent, 1941	29	79.3	89.3
Triatoma pintodiasi Jurberg, Cunha & Rocha, 2013	26	88.0	84.0
Triatoma platensis Neiva, 1913	28	74.1	71.4
Triatoma pseudomaculata Correa & Espínola, 1964	56	70.9	75.0
Triatoma rubrovaria (Blanchard, 1843)	55	59.3	79.6
Triatoma sherlocki Papa, Jurberg, Carcavallo, Cerqueira & Barata, 2002	32	93.5	90.6
Triatoma sordida (Stål, 1859)	96	81.2	92.7
Triatoma tibiamaculata (Pinto, 1926)	41	92.7	85.4
Triatoma vandae Carcavallo, Jurberg, Rocha, Galvão, Noireau & Lent, 2002	30	69.0	79.3
Triatoma vitticeps (Stål, 1859)	48	85.1	80.9
Triatoma williami Galvão, Souza & Lima, 1965	20	70.6	70.0
Total sample size and average success rates	1584	83.9	86.7

CHAPTER 2. Deep learning technology improves auditory biodiversity assessment and new candidate species identification in a hyper-diverse and yet underestimated species assemblage from an island archipelago

2.1 Introduction

Species new to science are continuously being described, and therefore many evolutionary, ecological, and behavioral phenomena and processes remain to be discovered (Scheffers *et al.* 2013; Tonini *et al.* 2020). However, habitat destruction is triggering the rapid loss of species unknown to science (Bryan *et al.* 2013; Tapley *et al.* 2018). Recent efforts to overcome this arms race between species discovery and extinction (Klein, McKown & Tershy 2015; González-del-Pliego *et al.* 2019) have focused on development of automated monitoring devices, such as passive recorders for monitoring species with acoustic mate-recognition signals (i.e., bats, birds, frogs, crickets; Chen & Wiens 2020). Research using conventional recordings and new automated devices, however, has generated exorbitant quantities of data, at a pace faster than they can be analyzed (Brabant *et al.* 2018). As such, a need for automated processing tools and acoustic species identification has emerged.

Patterns of seasonal phenology, diel activity, habitat use, and focal species monitoring have been the subject of acoustic signal inventories (Sugai *et al.* 2019) and biodiversity assessments (Wimmer *et al.* 2013). The use of advertisement calls for integrative amphibian species delimitation and identification has increased steadily (Vieites *et al.* 2009; Brown & Stuart 2012; Philippe, Felipe & Celio 2017), but automating species discovery from environmental recordings has not been applied widely.

The advertisement calls of amphibians are primary phenotypes for mate-recognition (Gerhardt 1994; Wells & Schwartz 2007) and analyses of temporal and spectral acoustic data have

been used widely to assign populations to species (Gerhardt 1978; Vignal & Kelley 2007; Feinberg *et al.* 2014). Once species-specific signals have been characterized quantitatively, automated classification methods can be employed for identification and assignment of species from natural soundscapes (Aide *et al.* 2013; Zhao *et al.* 2017). However, fundamental challenges arise when species new to science are recorded: classification is prevented by an absence of their temporal/spectral signal properties in training data sets.

Increased use of images for automated species identification (Rzanny *et al.* 2017; Villa, Salazar & Vargas 2017; Khalighifar *et al.* 2019) suggests that the highly stereotyped nature of anuran signals (Narins & Capranica 1977; Gerhardt 1994) as characterized in sound spectrograms (*sensu* Wells 2010), could be exploited for image-based species identification. To explore this possibility, we selected the Philippine frog genus *Platymantis* (family Ceratobatrachidae), for three reasons. First (A) *Platymantis* has been the focus of intensive surveys of advertisement call variation (Brown, Alcala & Diesmos 1997; Brown *et al.* 1997; Brown *et al.* 1999; Brown, Alcala & Diesmos 1999) and the fast pace of species description (Brown & Gonzalez 2007; Brown & Stuart 2012; Diesmos *et al.* 2015), and identification of candidate species for future taxonomic studies (Brown *et al.* 2015a,b) demonstrate the prevalence of considerable underestimated species diversity (Brown *et al.* 2013, 2015b). Additionally, (B) as the focus of recent molecular phylogenetic analyses, *Platymantis* is a demonstrably monophyletic, Philippine-endemic clade (Brown *et al.* 2015b), sister to the similarly-diverse Papuan genus *Cornufer*. Thus, the geographic and systematic understanding of *Platymantis* is much improved over earlier work (Inger 1954). Third, (C) calls of nearly all recognized *Platymantis* species (AmphibiaWeb 2020, https://amphibiaweb.org) are available in the public domain. The combination of available call resources (large samples of calls from recognized and undescribed species), and a robust

phylogeny as an historical framework render *Platymantis* a promising focal system for research in automating species identification, with intended future expansion of the system to incorporate comparative phylogenetic methods.

Automation offers a solution for tasks requiring repetition when experienced workers are lacking or cost-prohibitive (Gaston & O'Neill 2004). Recent computer science developments in image classification and signal processing provide new tools that may improve biodiversity assessment in taxa representing visual identification challenges (MacLeod, Benfield & Culverhouse 2010; Guirado *et al.* 2019). Although automated species identification systems using handcrafted feature extraction has shown promising results (Holmgren, Persson & Söderman 2008; Kumar *et al.* 2012; Gurgel-Gonçalves *et al.* 2017), it requires advanced expertise and programming to accomplish robust performance.

In recent years, a robust group of classifiers has been introduced (Deep Neural Networks, DNNs; Schmidhuber 2015), which outperform existing methods in various classification tasks (Ramcharan *et al.* 2017; Smith *et al.* 2019). One of the state-of-the-art DNN platforms is TensorFlow (Abadi *et al.* 2016), an open-source software platform designed by the Google Brain Team (https://www.research.google.com/teams/brain). One crucial advantage of applying TensorFlow is the Transfer Learning technique, which is a shortcut for achieving high-performance classification. This approach involves using a large dataset to train a model, and then re-training with a new calibrating dataset both to improve identification rates at lower computational cost.

Here, we applied TensorFlow Inception v3 (Szegedy *et al.* 2016), implemented in the Linux environment (Ubuntu, version 18.04; https://www.ubuntu.com), to explore the challenge of automating frog species identification. We explored two major challenges in this study: (1)

whether TensorFlow is able to discriminate among species of Philippine forest frogs, based on simple visual-image representations of auditory signals, and (2) whether TensorFlow could go beyond simple inventory to accelerate new candidate species discovery by objectively identifying undescribed species. To address these questions, we designed an automated, DNN-based species identification system for 41 described *versus* undescribed *Platymantis* species, which we tested using single-note call spectrograms. This study thereby lays a foundation for automated identification capabilities in biodiversity conservation and assessment via auditory signals.

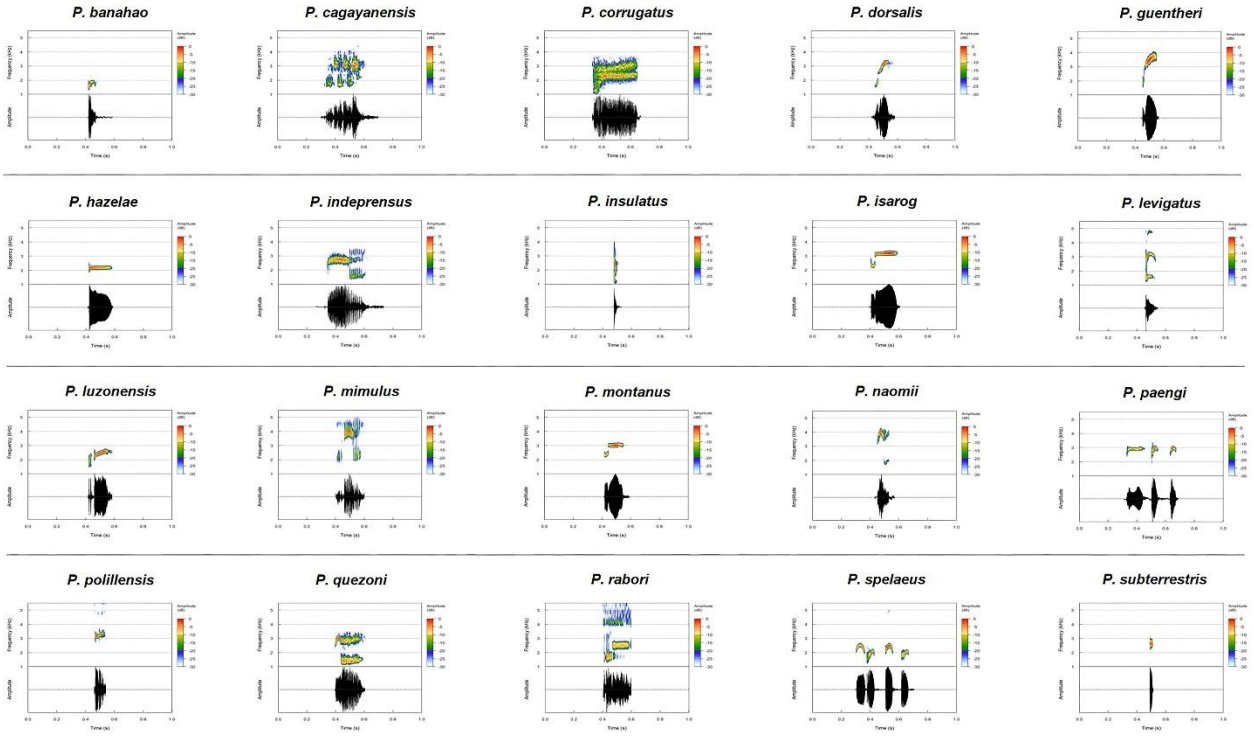

Figure 1. Examples of images used for identification of 20 currently recognized species of Philippine forest frogs, genus *Platymantis* (Brown *et al.* 2015b), available from the Cornell Lab of Ornithology Macaulay Library. Each spectrogram has the same time duration (one second), and frequency limits (1-5.5 kHz).

2.2 Materials and Methods

2.2.1 Data processing

We obtained and analyzed frog recordings from two sources: (1) a large collection of Philippine frog advertisement calls collected, archived (by RMB and colleagues), and made publicly available via Cornell University's Laboratory of Ornithology and Macaulay Library of Natural Sounds (https://www.macaulaylibrary.org), and (2) recent collections (2005–2019) of numerous undescribed species (review: Brown *et al.* 2015b; Diesmos *et al.* 2015), to augment sample sizes of previously described species and add distinctive new candidate species identified with genetic and phenotypic characters (RMB, unpublished data). Additional collection-associated natural history information, frog microhabitats, community composition, recording methodology (device information, digitization specifications), and metadata are available via the Macaulay Library portal and the KU Herpetology online Specify database, as well as via GBIF, iDigBio, and other aggregators; behavioral context of calls, and qualitative descriptions of calls are available in original descriptions (e.g., Brown & Gonzalez 2007; Siler *et al.* 2007, 2010; Brown *et al.* 2015a).

We surveyed 175 recordings, representing 20 species (Figure 1) using the cross-platform audio editor Ocenaudio (https://www.ocenaudio.com). This software is based on Ocen Framework, a powerful library to simplify and standardize manipulation and analysis of audio files. We clipped 20 high-quality single notes per each species, and saved each as 32-bit, single-channel WAV files (44.1 kHz sampling rate). To standardize temporal scale across comparisons, we designated a duration of one second; all known *Platymantis* species' single notes fit this range. To do so, we added silence in equal length to the beginning and the end of each clipped single note. Then, we used R packages warbleR (Araya-Salas & Smith-Vidaurre 2017) and Seewave (Sueur, Aubin & Simonis 2008) to generate spectrograms across a standardized range of

frequencies, 1.0–5.5 kHz; all known *Platymantis* calls fall within this range (Figure 1). To generate oscillograms, we chose a fast-Fourier transformation (FFT) of 512 points, with 90% overlap between two successive windows. We saved all spectrograms as Portable Network Graphics (PNGs).

2.2.2 Model architecture

Convolutional neural networks (CNNs) are a subset of DNNs that are specialized for image classification tasks and pattern recognition. One of the main advantages of CNNs is the ability to perform automated feature extraction, eliminating the need for hand-crafted feature extraction. CNN architecture is built on three types of layers: (1) convolutional layers, which are the most important because they apply hierarchical feature extraction and decomposition of input images; (2) pooling layers, which carry out operations to reduce numbers of parameters and necessary computation; and (3) fully connected layers, which perform the actual classification at the end of the pipeline.

CNNs require large training datasets to achieve accurate classification rates. Although training on a large dataset provides a powerful framework, building and training a CNN from scratch is both computationally expensive and time consuming. To overcome these limitations, we used a transfer learning technique. Transfer learning means using experience acquired from classification task A in classification task B. This technique allows the user to retrain the final layer of an existing model on the training set associated with a new classification task. One of the most successfully implemented models of transfer learning is Inception v3 (Szegedy *et al.* 2016)—a CNN, implemented in TensorFlow (Abadi *et al.* 2016). This CNN consists of 48 layers, and is trained on >1M images from the ImageNet database (http://www.image-net.org). Inception v3 is

widely recognized for outperforming other models in challenges involving classifying images into thousands of classes (Russakovsky *et al.* 2015).

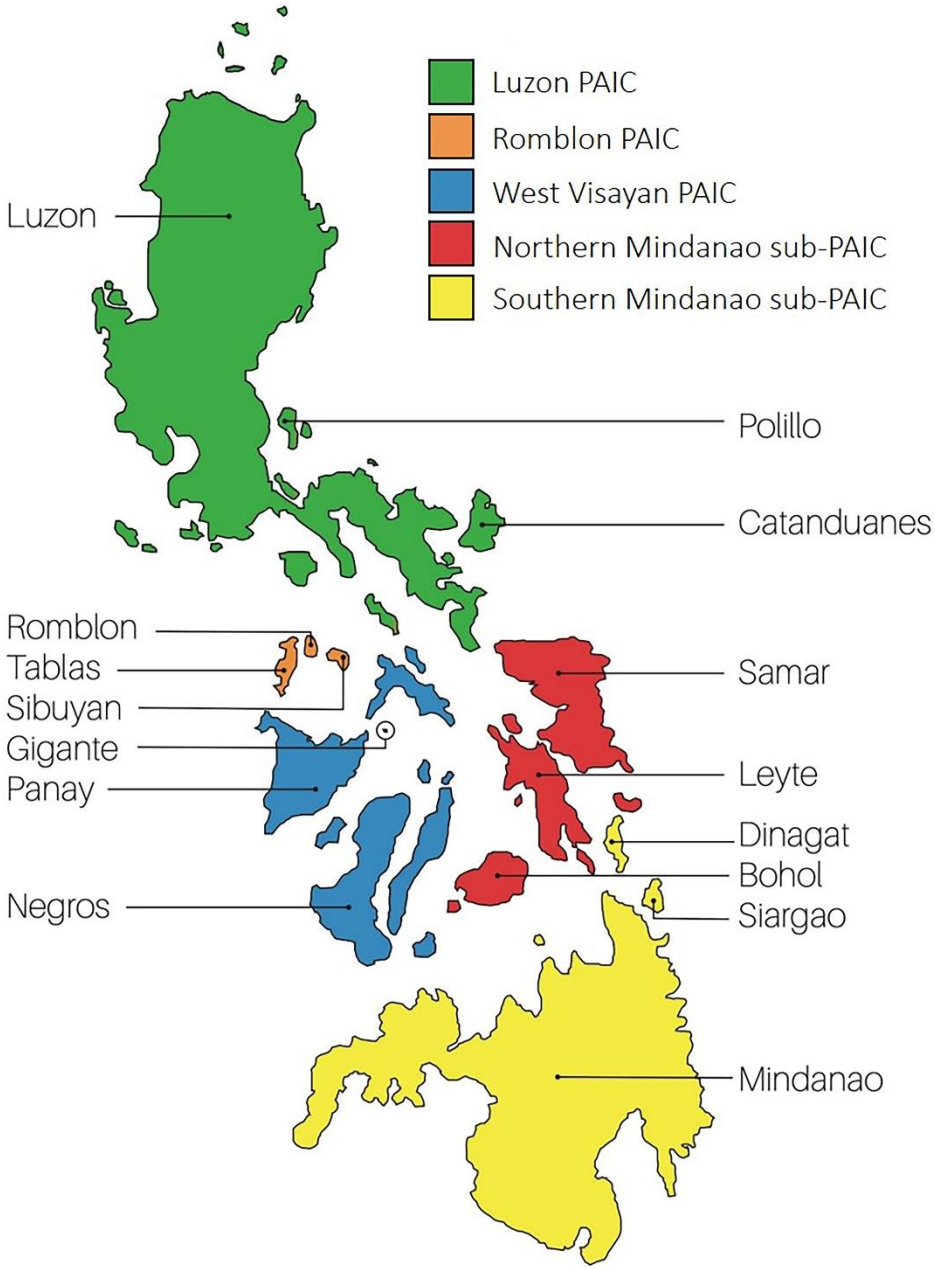

Figure 2. Map of Philippines, with Pleistocene Aggregate Island Complex (PAICs) faunal regions (colored shading) used to create realistic species pools (Challenge 4) to enhance identification. Map simplified to include only PAICs relevant to this study.

2.2.3 Classification challenges

We designed four classification challenges using single-note spectrograms and we assessed whether TensorFlow (Inception v3) is capable of successfully identifying *Platymantis* species based on frequency distributions of individual call notes of each species. For the training process, we modified two parameters of the model: (1) number of training steps, and (2) validation percentage. We explored different numbers of training steps and compared results to find an optimum balance between computing time and classification efficiency. Given our limited number of images per species (n=20), we increased the validation parameter to 20%. For the same reason, we used a leave-one-out cross validation technique (Molinaro, Simon & Pfeiffer 2005) to evaluate model performance in Challenges 1, 3, and 4. The four classification challenges we explored are as follows:

1. Applying TensorFlow to identify species available from the Cornell Library of Natural Sounds: We applied TensorFlow to data from the Macaulay Library, including 20 recognized species of *Platymantis* (from among 33 described forms; Brown *et al.* 2015b), as classes for input. We generated 400 spectrograms (20 per species) for the identification challenge.

2. Challenging TensorFlow with species not in the reference library: We trained TensorFlow on all images (i.e., 400 spectrograms) from Challenge 1 as an image reference library. Then, we applied the trained model to a test dataset from 22 robustly-identified species from recent field surveys by RMB. We addressed two questions: (1) could TensorFlow identify species existing in its reference library among the unknown species when they constitute new recordings obtained from different individuals? And, more importantly, (2) how does TensorFlow perform when it encounters species that do not exist in its reference library? To answer these questions, we

generated 20 spectrograms per species for this new dataset, resulting in a total of 440 spectrograms, to be subjected to identification using the CNN developed in Challenge 1.

3. Challenging TensorFlow to identify all 41 species for which recordings are available: We increased the number of species in the TensorFlow reference library to 41 by adding 21 new, undescribed species (one species in the new dataset was already present in reference library, owing to a recent taxonomic change; Brown *et al.* 2015b). Then, we followed the same procedure as in Challenge 1, to test performance on a reference library that is twice as large as the original (i.e., 41 classes).

4. Faunal region-based identification in natural species pools: We used distributional data (Brown *et al.* 2015b; Diesmos *et al.* 2015) to create subsets of species, with the goal of generating separate classification tasks with lower numbers of classes (species) per task (local species pools, reflecting documented patterns of co-distributed species from the archipelago's faunal regions). First, we grouped the 41 species (20 described species, plus 21 undescribed candidate species) based on Philippine islands they inhabit, resulting in 15 subsets of co-occurring species in "communities" of 3 to 27 species (Figure 2). Then, we trained and tested the classifier employing the set of samples from the species found on those islands. Finally, we calculated the overall correct identification rate across 15 islands to compare with that based on the full reference library.

2.3 Results

We calibrated models for classification challenges using different numbers of training steps, and considered two factors to find optima for training TensorFlow (Khalighifar *et al.* 2019): correct identification rate and processing time. As a result, for all challenges except Challenge 3,

we chose 4000 training steps as an optimum number. For Challenge 3, given the number of species (41 species), 8000 training steps proved to be the optimum number. The details of results associated with each classification task are as follows:

Species Name	*P. banahao*	*P. cagayanensis*	*P. corrugatus*	*P. dorsalis*	*P. guentheri*	*P. hazelae*	*P. indeprensus*	*P. insulatus*	*P. isarog*	*P. levigatus*	*P. luzonensis*	*P. mimulus*	*P. montanus*	*P. naomii*	*P. paengi*	*P. polillensis*	*P. quezoni*	*P. rabori*	*P. spelaeus*	*P. subterrestris*
P. banahao	1.00																			
P. cagayanensis		1.00																		
P. corrugatus			1.00																	
P. dorsalis				1.00																
P. guentheri				0.1	0.90															
P. hazelae						0.95					0.05									
P. indeprensus		0.05					0.95													
P. insulatus								1.00												
P. isarog						0.1			0.70				0.2							
P. levigatus										1.00										
P. luzonensis											0.95								0.05	
P. mimulus												1.00								
P. montanus	0.05								0.1				0.75			0.05			0.05	
P. naomii						0.05								0.95						
P. paengi															1.00					
P. polillensis																1.00				
P. quezoni			0.05				0.05										0.80	0.1		
P. rabori																		1.00		
P. spelaeus											0.05		0.05						0.90	
P. subterrestris																				1.00

Figure 3. Confusion matrix for 20 currently recognized species of Philippine forest frogs, genus *Platymantis* (*sensu* Brown *et al.* 2015b) using a leave-one-out cross-validation technique. Red = correct identification; yellow = misidentifications. All values of zero are removed for ease of visualization.

Challenge 1. We created a confusion matrix to depict TensorFlow's initial results with 20 species (Figure 3). The overall correct identification rate was 94.3%. We achieved 100% correct identification rate for 11, and 90% or above for 17, species. The lowest identification rates were for closely-related species *Platymantis isarog* and *P. montanus*, with 70 and 75% correct

classification, respectively. We did not detect any systematic errors in TensorFlow classification, such as repeatedly confusing one species with another.

Species Name	*P. banahao*	*P. cagayanensis*	*P. corrugatus*	*P. dorsalis*	*P. guentheri*	*P. hazelae*	*P. indeprensus*	*P. insulatus*	*P. isarog*	*P. levigatus*	*P. luzonensis*	*P. mimulus*	*P. montanus*	*P. naomi*	*P. paengi*	*P. polillensis*	*P. quezoni*	*P. rabori*	*P. spelaeus*	*P. subterrestris*	Certainty
P. biak (formerly sp. 37)	8 53.4%				1 34.1%	7 53.4%			1 37.4%	2 51.5%				1 59.8%							51.7%
P. cf *cornutus*				6 44.5%				5 53.8%		2 62.0%				1 85.2%						6 64.0%	56.5%
P. cf *lawtoni*			1 36.9%			2 45.8%	2 67.7%		13 64.6%								2 25.3%				57.7%
P. cf *sierramadrensis*					1 88.5%				17 75.5%				2 72.0%								75.8%
P. diesmosi (formerly sp. 2)								20 94.3%													94.3%
P. isarog									20 98.1%												98.1%
P. sp. 3					4 47.7%				11 76.7%				5 65.1%								68.0%
P. sp. 4					2 76.4%				17 91.1%				1 54.7%								87.8%
P. sp. 6				1 72.8%	3 48.6%						2 50.6%			14 75.7%							69.0%
P. sp. 8				3 64.7%	1 42.4%						6 60.7%			7 69.8%		3 56.1%					62.9%
P. sp. 12					2 40.8%				2 32.7%	1 32.3%	1 29.0%	2 49.5%	2 29.9%			4 55.0%	2 42.5%	1 21.4%	3 75.3%		38.4%
P. sp. 14		10 72.2%													10 70.8%						71.5%
P. sp. 15		8 60.9%													10 54.4%		1 25.3%		1 32.7%		54.5%
P. sp. 16					9 74.0%										10 92.9%				1 77.6%		83.6%
P. sp. 17								6 62.1%		3 74.1%										11 72.3%	69.5%
P. sp. 22				1 28.3%	11 53.3%								1 41.5%	7 62.8%							54.8%
P. sp. 23					16 70.0%				1 36.5%			1 42.8%		1 47.6%		1 35.6%					64.1%
P. sp. 25			12 73.6%		8 90.9%																80.5%
P. sp. 38 "double call"		2 41.4%	1 48.2%		3 37.8%			1 53.6%						4 58.4%	1 33.0%				8 61.4%		52.8%
P. sp. 38 "single call"					17 73.5%								3 73.5%								73.5%
P. sp. 44					10 53.1%				2 43.9%			2 49.9%		1 23.7%		1 63.1%	2 47.5%	2 50.9%			50.1%
P. sp. 49				6 56.7%	1 39.7%			8 66.5%												5 56.2%	59.6%

Figure 4. Confusion matrix resulting from challenging TensorFlow with potential species unknown to the reference library—numeric species identifiers from Brown *et al.* (2015b). Columns = species in the reference library; rows = potential unknown species, with exception of *P. isarog* (see text). Red numbers = certainty rates below 40%; black = 41–85%; green >85%. Far right column = average certainty rate for species identifications.

Challenge 2. TensorFlow provides each image identification task with two elements/features: suggested species names and a certainty rate. Certainty rate can be a factor by which to evaluate classifier performance on test images as well. After applying TensorFlow on a testing dataset consisting of 22 species, we considered two factors to evaluate model performance:

(1) number of images per class assigned to a species present in reference library, and (2) average certainty rate associated with those identifications (Figure 4). Among the 22 species in the test dataset, only one, *P. isarog*, was also present in the reference library; however, the remainder were new to the training set. As a result, it was impossible for TensorFlow to provide a correct answer for the other 21 species. The overall certainty rate for those 21 species was 65.5%. However, in Challenge 1, the overall certainty rate for the 20 species present in reference library was 83.6%. After using Mann-Whitney U test in Python 3.8.2, the results showed that TensorFlow yielded a significantly lower certainty rate for the 21 new species to the reference library (U=56379.5, *P*-value=1.09e-19).

Certainty rates in Challenge 2 ranged from 38.4–98.1%. TensorFlow yielded an overall certainty rate ≥90% for only 2 of the 22 species in our test dataset, one of which was a correct identification. That is, all images associated with *P. isarog* identified as *P. isarog* with an average certainty rate of 98.1%. In contrast, all *P. diesmosi* images were identified as *P. insulatus* with a 94.3% average certainty rate. Another species with a relatively high average certainty rate was *P.* sp. 4, a taxon originally described as "*P. rivularus*" (Taylor 1923), which is expected to be elevated from synonymy of *P. subterrestris* with ongoing studies (Brown *et al.* 2015b). Individuals of this population were identified as *P. isarog*, *P. guentheri*, and *P. montanus*, with an average certainty rate of 87.8%. The lowest certainty rates yielded by TensorFlow were for three undescribed species, *P.* sp. 12 (onomatopoeically nicknamed "churink" with a 38.4% average certainty rate), *P.* sp. 44 ("Ee-yow" with a 50.1% average certainty rate), and *P. biak* (with a 51.7% average certainty rate), which were classified as 10, 7, and 6 different species, respectively. Among the species in the reference library, the most frequently suggested species was *P. guentheri*, which was suggested for 15 of 22 species in our test dataset. However, regarding two evaluation factors

mentioned above, in none of those cases, could *P. guentheri* be considered as the primary identification for those species (Figure 4).

Table 1. Summary of species analyzed, success in terms of number of images correctly identified, failure in terms of number of misidentifications, and correct identification rate for 20 described species of *Platymantis* (*sensu* Brown *et al.* 2015b) and 21 new, undescribed candidate species. Sample size for each species is 20 images; undescribed species specifier numbers follow Brown *et al.* (2015b).

Species	Success	Failure	Correct identification rate
Platymantis banahao	20	0	1.00
P. cagayanensis	20	0	1.00
P. corrugatus	20	0	1.00
P. dorsalis	20	0	1.00
P. guentheri	18	2	0.90
P. hazelae	19	1	0.95
P. indeprensus	19	1	0.95
P. insulatus	20	0	1.00
P. isarog	14	6	0.70
P. levigatus	20	0	1.00
P. luzonensis	19	1	0.95
P. mimulus	20	0	1.00
P. montanus	16	4	0.80
P. naomii	19	1	0.95
P. paengi	20	0	1.00
P. polillensis	20	0	1.00
P. quezoni	16	4	0.80
P. rabori	20	0	1.00
P. spelaeus	18	2	0.90
P. subterrestris	20	0	1.00
P. biak (formerly sp. 37)	20	0	1.00
P. cf *cornutus*	19	1	0.95
P. cf *lawtoni*	18	2	0.90
P. cf *sierramadrensis*	18	2	0.90

Species	Success	Failure	Correct identification rate
P. diesmosi (formerly sp. 2)	20	0	1.00
P. sp. 3	14	6	0.70
P. sp. 4	18	2	0.90
P. sp. 6	18	2	0.90
P. sp. 8	19	1	0.95
P. sp. 12	19	1	0.95
P. sp. 14	20	0	1.00
P. sp. 15	20	0	1.00
P. sp. 16	18	2	0.90
P. sp. 17	19	1	0.95
P. sp. 22	20	0	1.00
P. sp. 23	19	1	0.95
P. sp. 25	19	1	0.95
P. sp. 38 “double call”	19	1	0.95
P. sp. 38 “single call”	20	0	1.00
P. sp. 44	19	1	0.95
P. sp. 49	18	2	0.90

Challenge 3. Training TensorFlow on all 41 species, we observed a mere 0.2% decline in overall correct identification rate (i.e., overall correct identification rate for 41 species was 94.1 *versus* 94.3% for 20 species in Challenge 1). Although we added 21 species to the reference library, we observed no negative impact on correct identification rates, even though such impacts were noted in our previous work (Khalighifar *et al.* 2019). TensorFlow was able to identify 37 of 41 species with ≥90% correct identification rate; 17 species were identified with 100% identification rate. The lowest correct identification rates were for *P. isarog* and *P.* sp. 3 (*sensu* Brown *et al.* 2015b) with 70% (Table 1). Similar to Challenge 1, no species was detected to be mis-identified repeatedly as another species in particular.

Challenge 4. Classifier accuracy generally improves as the number of classes that must be distinguished decreases (Khalighifar *et al.* 2019). To improve accuracy and reduce numbers of species, we used geographic information (species distributions from Diesmos *et al.* 2015) to refine identification efforts. Species subsets ranged from 3 on Gigante, Panay, and Romblon to 27 on Luzon (Figure 5). As expected, TensorFlow was able to increase the overall identification rate from 94.1 to 98.7% by incorporating distributional information.

Average identification rates for more than half of the islands (8 of 15) were 100%. Luzon Island hosts 27 species, and scored the lowest overall identification rate, 94.6%. We noticed that another factor affecting overall identification rate for each island was species composition. That is, for some islands with lower numbers of species, we nonetheless found a lower overall identification rate as well. For example, species from Siargao and Dinagat islands, with the same species composition (*Platymantis corrugatus*, *P. dorsalis*, *P. guentheri*, and *P. rabori*) had a lower overall correct identification rate than Mindanao Island, a far larger landmass with at least six species (96.3% *versus* 100%).

2.4 Discussion

Pressures such as habitat loss, invasive species, pollution, and climate change are increasingly affecting amphibian populations (Blaustein *et al.* 2011; Pili *et al.* 2019). Frog populations have been regarded as excellent bio-indicators owing to their sensitivity to environmental change (Katti, Ghodgeri & Goundadkar 2016; Houston, Melzer & Black 2018). The Philippines has impressive endemic biodiversity (~110 amphibian species; 90% endemic; Diesmos *et al.* 2015), and documentation of species diversity using rapid, efficient inventory techniques, capable of detecting new species, is increasingly needed. This challenge is the focus

of work which we hope will automate rapid candidate species discovery, to then allow researchers to target strategically subsequent statistical species validation (e.g., Barley *et al.* 2013; Chan *et al.* 2017), to be followed by formal taxonomic description (Brown et al. 2015b) as logical steps towards characterizing the archipelago's amphibian biodiversity.

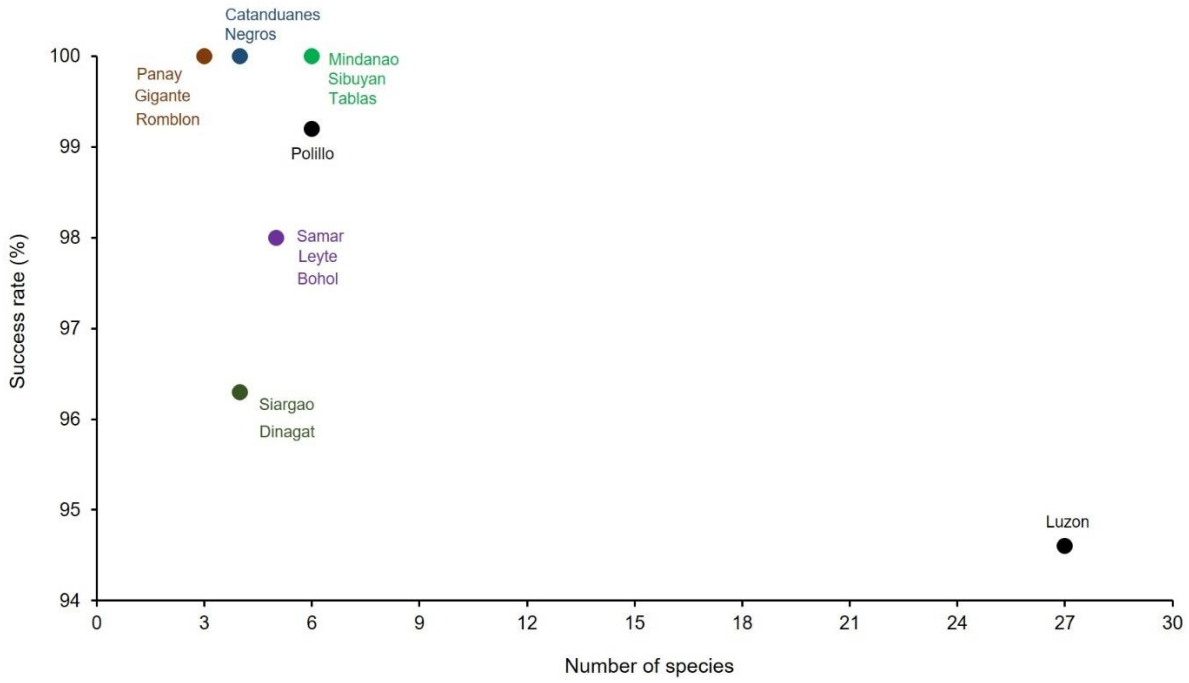

Figure 5. Island-based identification for 20 currently-recognized species of Philippine forest frogs (*sensu* Brown *et al.* 2015b) and 21 new, undescribed *Platymantis* species, incorporating biogeographic information (island bank-based faunal regions; Brown *et al.* 2013) and species distribution information (Diesmos *et al.* 2015). Note that although we incorporated distributional information (PAICs) in our analyses, this figure summarizes individual island results. Each point is only associated with two elements; (1) number of species per island (*x*-axis), and (2) correct identification rates (*y*-axis).

To this end, we explored CNN Inception v3 as a means by which to automate several key inferential tasks. We successfully identified *Platymantis* species with an impressive overall identification rate based on single call note characteristics, which was a surprising result for field biologists (RMB and colleagues) who are accustomed to discovering and describing *Platymantis*

species based on temporal patterns of note repetition, complex call elements, and rich spectral properties of many calls (Brown & Gonzalez 2007; Siler *et al.* 2007, 2010; Brown *et al.* 2015b). Incorporating distributional data (grouping species by Pleistocene island bank-based faunal regions) allowed us to create realistic subsets of species pools. This further increased overall identification rates by reference to the relevant, naturally co-occurring species pool.

2.4.1 Deep neural networks

TensorFlow's flexible architecture allows easy adaptation and deployment on different platforms including desktops, clusters of servers, edge devices, and mobile phones (https://www.tensorflow.org/). Inception v3 is a sophisticated network given the number of layers (48), and is already trained on more than a million images.

However, one major challenge was the number of call notes (input spectrograms) available for each species. To address this, future studies should access more recordings per species, particularly those represented now by recordings of few individuals. Despite this limitation, our study is a novel use of a deep-learning platform to distinguish between closely-related species of frogs using simple, single-note, two-dimensional depictions of primary mate-recognition cues (mating calls). A second challenge, also related to sample size, was model validation. We would have been more satisfied with our evaluation if we had more recordings from numerous individuals to utilize *k*-fold cross-validation (Kohavi 1995) instead of leave-one-out approaches, because the former offers greater independence between calibration and evaluation datasets.

2.4.2 Taxonomic identification

In this study, we demonstrate the efficacy of deep learning technology for reliably identifying—and distinguishing among—closely-related frog species, as exemplified by single-note call segments (Figure 1). Given known phylogenetic relationships (Brown *et al.* 2015b), we were surprised that multiple closely-related species pairs were distinguished from one another perfectly (100% success). For example, species pairs *P. indeprensus* and *P. mimulus* (both members of the subgenus *Lupacolus*), *P. hazelae* and *P. montanus* (subgenus *Tirahanulap*), and *P. levigatus* and *P. insulatus* (subgenus *Lahatnanguri*) could each be predicted, by virtue of their close phylogenetic relationships, to have similar spectral (frequency-related) and temporal (time-related) call properties—which they do (Figure 1). Still, with only a single isolated note per species, TensorFlow is able to distinguish them and correctly classify species' identity, when presented with a large sample of positively identified individual notes (i.e., known populations of confidently-identified species, based on fully documented voucher specimens deposited in biodiversity repositories).

That said, illustrative examples of how the methods failed in our study—cases where identification was problematic, attributed to multiple species, or when a sample of notes were classified to wrong species—are worthy of consideration. In these cases (Table 1), two categories of identification errors emerge: Type A stems from closely-related species, with brief, pure-tone, constant-frequency calls, whose calls are exceptionally simple, intra-specifically invariant, and even inter-specifically quite difficult to distinguish. Referred to as "cloud frogs," members of subgenus *Tirahanulap* (former "*P. hazelae* Group" species; Brown *et al.* 2015b; Diesmos *et al.* 2015) are all diminutive (1–3 g body mass), primarily higher-elevation moist, closed-canopy shrub frogs. Their close phylogenetic relationships (Brown *et al.* 2015b) and remarkably similar

microhabitat preferences render it no surprise to us that TensorFlow had difficulty distinguishing *P. isarog, P. montanus, P.* sp. 3, *P. sierramadrensis,* and *P. lawtoni.* Another category of error was exemplified (Type B) by instances of apparent convergence in frequency modulation, exemplified by unrelated species such as *P. dorsalis* (subgenus *Lupacolus*), *P. guentheri* (*Tahananpuno*), and the amplitude-modulated, rapidly-repetitive pulse-train calls of taxa like *P. luzonensis, P.* sp. 6, and *P.* sp. 8. (*Tahananpuno*). In these taxa, it is little surprise that single-note call components are occasionally mis-specified by TensorFlow, given that they are essentially homologous call elements (Brown *et al.* 2015b), temporally arranged to differ only by numbers of notes per call and calling rate (Brown *et al.* 1997a,b; Brown *et al.* 1999a,b; Brown *et al.* 2015b). However, given that only a few closely-related species pairs exhibit overlapping, sympatric geographic ranges (Diesmos *et al.* 2015), our confidence is further bolstered with the confirmation that automated discrimination can be enhanced by limitation of species classes to realistic species pools (Brown *et al.* 2013; see below). In summary, we take these results as encouraging in that the efficacy of automated species identification can be improved with biogeographic information.

Our refinement of the method, using biogeographically-relevant species sets and limiting species identifications to co-distributed taxa, resulted in a dramatic improvement in method performance (Figure 4), particularly when considering caveats discussed above. By limiting the possible universe of a species' identification to the biogeographically-relevant species pools, i.e., we both (1) improved performance of identifications of known taxa, and (2) drew attention to (analytically singled out) unknown, new, or undescribed taxa (Figures 2, 3). These features will be valuable in identifying taxa for subsequent 'validation' of unconfirmed candidate species, using independent data streams (phenotypic data, genetic information, ecological characteristics, etc.).

Across broader taxonomic scales and phylogenetic relationships (e.g., Chan & Brown 2017), other means (biogeographic realm, ectomorph type, classification, etc.) of restricting/limiting candidate species pools may prove useful for 'fine-tuning' of TensorFlow's automation of species recognition. Additional caveats for future consideration include (1) single notes per species and (2) sample sizes, which will be limited for rare species, those that occur at naturally low abundances, or taxa characterized by reduced detection probabilities due to cryptic microhabitat preferences, narrow activity patterns, or seasonally-limited reproductive cycles (Wells 2010). Avenues for future development of these methods in our immediate plans include application to additional taxonomic groups (e.g., insects, birds), and automation of call detection from environmental sound samples as a precursor step to automated species identification.

2.5 Acknowledgements

Philippine *Platymantis* frog calls were collected with support from the U. S. National Science Foundation's former Doctoral Dissertation Improvement Grant (DEB 0073199; 2001–2003) and a Biotic Surveys and Inventories grant (DEB 0743491; 2008–2012). Collection and co-curated voucher specimens and their associated digital media specimens, archival digitization, data verification, and online serving of digital media specimens was made possible by an NSF Thematic Collections Network (TCN) program grant (DEB 1304585; 2013–2018). Recent extended specimen collection and curation has been supported by NSF DEB 1654388 and 1557053, with further support from the KU Biodiversity Institute's Rudkin Research Exploration (REX) Fund and KU College of Liberal Arts and Sciences Docking Scholar Fund. We thank A. Diesmos, J. Fernandez, C. Siler, and C. Meneses for help recording frogs.

2.6 References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G. & Isard, M. (2016) Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation*, pp. 265-283. Savannah, USA.

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G. & Alvarez, R. (2013) Real-time bioacoustics monitoring and automated species identification. *PeerJ,* **1,** e103.

Araya-Salas, M. & Smith-Vidaurre, G. (2017) warbleR: An R package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution,* **8,** 184-191.

Barley, A.J., White, J., Diesmos, A.C. & Brown, R.M. (2013) The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. *Evolution,* **67,** 3556-3572.

Blaustein, A.R., Han, B.A., Relyea, R.A., Johnson, P.T., Buck, J.C., Gervasi, S.S. & Kats, L.B. (2011) The complexity of amphibian population declines: Understanding the role of cofactors in driving amphibian losses. *Annals of the New York Academy of Sciences,* **1223,** 108-119.

Brabant, R., Laurent, Y., Dolap, U., Degraer, S. & Poerink, B.J. (2018) Comparing the results of four widely used automated bat identification software programs to identify nine bat species in coastal Western Europe. *Belgian Journal of Zoology,* **148,** 119-128.

Brown, R. & Stuart, B. (2012) Patterns of biodiversity discovery through time: An historical analysis of amphibian species discoveries in the Southeast Asian mainland and island archipelagos. *Biotic Evolution and Environmental Change in Southeast Asia***,** 348-389.

Brown, R.M., De Layola, L.A., Lorenzo, A., Diesmos, M.L.L. & Diesmos, A.C. (2015a) A new species of limestone karst inhabiting forest frog, genus *Platymantis* (Amphibia: Anura: Ceratobatrachidae: subgenus *Lupacolus*) from southern Luzon Island, Philippines. *Zootaxa,* **4048,** 191-210.

Brown, R.M. & Gonzalez, J.C. (2007) A new forest frog of the genus *Platymantis* (Amphibia: Anura: Ranidae) from the Bicol Peninsula of Luzon Island, Philippines. *Copeia,* **2007,** 251-266.

Brown, R.M., Siler, C.D., Oliveros, C.H., Esselstyn, J.A., Diesmos, A.C., Hosner, P.A., Linkem, C.W., Barley, A.J., Oaks, J.R. & Sanguila, M.B. (2013) Evolutionary processes of diversification in a model island archipelago. *Annual Review of Ecology, Evolution, and Systematics,* **44,** 411-435.

Brown, R.M., Siler, C.D., Richards, S.J., Diesmos, A.C. & Cannatella, D.C. (2015b) Multilocus phylogeny and a new classification for Southeast Asian and Melanesian forest frogs (family Ceratobatrachidae). *Zoological Journal of the Linnean Society,* **174,** 130-168.

Brown, W., Alcala, A., Ong, P. & Diesmos, A. (1999) A new species of *Platymantis* (Amphibia: Ranidae) from the Sierra Madre Mountains, Luzon Island, Philippines. *Proceedings of the Biological Society of Washington,* **112,** 510-514.

Brown, W.C., Alcala, A. & Diesmos, A.C. (1997) A new species of the genus *Platymantis* (Amphibia: Ranidae) from Luzon Island, Philippines. *Proceedings of the Biological Society of Washington,* **110,** 18-23.

Brown, W.C., Alcala, A., Diesmos, A.C. & Alcala, E. (1997) Species of the *guentheri* group of *Platymantis* (Amphibia: Ranidae) from the Philippines, with descriptions of four new species. *Proceedings of the California Academy of Sciences,* **50,** 1-20.

Brown, W.C., Alcala, A.C. & Diesmos, A.C. (1999) Four new species of the genus *Platymantis* (Amphibia: Ranidae) from Luzon Island, Philippines. *Proceedings of the California Academy of Sciences,* **51,** 449-460.

Bryan, J.E., Shearman, P.L., Asner, G.P., Knapp, D.E., Aoro, G. & Lokes, B. (2013) Extreme differences in forest degradation in Borneo: Comparing practices in Sarawak, Sabah, and Brunei. *PLoS ONE,* **8,** e69679.

Chan, K.O., Alexander, A.M., Grismer, L.L., Su, Y.C., Grismer, J.L., Quah, E.S. & Brown, R.M. (2017) Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian torrent frogs. *Molecular Ecology,* **26,** 5435-5450.

Chan, K.O. & Brown, R.M. (2017) Did true frogs 'dispersify'? *Biology Letters,* **13,** 20170299.

Chen, Z. & Wiens, J.J. (2020) The origins of acoustic communication in vertebrates. *Nature Communications,* **11,** 1-8.

Diesmos, A.C., Watters, J.L., Huron, N.A., Davis, D.R., Alcala, A.C., Crombie, R.I., Afuang, L.E., Gee-Das, G., Sison, R.V. & Sanguila, M.B. (2015) Amphibians of the Philippines, part I: Checklist of the species. *Proceedings of the California Academy of Sciences,* **62,** 457-539.

Feinberg, J.A., Newman, C.E., Watkins-Colwell, G.J., Schlesinger, M.D., Zarate, B., Curry, B.R., Shaffer, H.B. & Burger, J. (2014) Cryptic diversity in metropolis: Confirmation of a new leopard frog species (Anura: Ranidae) from New York City and surrounding Atlantic coast regions. *PLoS ONE,* **9,** e108213.

Gaston, K.J. & O'Neill, M.A. (2004) Automated species identification: Why not? *Philosophical Transactions of the Royal Society of London B,* **359,** 655-667.

Gerhardt, H.C. (1978) Mating call recognition in the green treefrog (*Hyla cinerea*): The significance of some fine-temporal properties. *Journal of Experimental Biology,* **74,** 59-73.

Gerhardt, H.C. (1994) The evolution of vocalization in frogs and toads. *Annual Review of Ecology and Systematics,* **25,** 293-324.

González-del-Pliego, P., Freckleton, R.P., Edwards, D.P., Koo, M.S., Scheffers, B.R., Pyron, R.A. & Jetz, W. (2019) Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians. *Current Biology,* **29,** 1557-1563. e1553.

Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D. & Herrera, F. (2019) Whale counting in satellite and aerial images with deep learning. *Scientific Reports,* **9,** 14259.

Gurgel-Gonçalves, R., Komp, E., Campbell, L.P., Khalighifar, A., Mellenbruch, J., Mendonça, V.J., Owens, H.L., de la Cruz Felix, K., 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.

Holmgren, J., Persson, Å. & Söderman, U. (2008) Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing,* **29,** 1537-1552.

Houston, W.A., Melzer, A. & Black, R.L. (2018) Recovery of reptile, amphibian and mammal assemblages in Australian post-mining landscapes following open-cut coal mining. *Proceedings of the Royal Society of Queensland,* **123,** 31-47.

Inger, R.F. (1954) Systematics and zoogeography of Philippine Amphibia. *Fieldiana Zoology,* **33,** 181-531.

Katti, P.A., Ghodgeri, M.G. & Goundadkar, B.B. (2016) Amphibian (*Euphlyctis cyanophlyctis*) *in vitro* ovarian culture system to assess impact of aquatic agrochemical contaminants on female reproduction. *Drug and Chemical Toxicology,* **39,** 104-110.

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,** 1404-1410.

Klein, D.J., McKown, M.W. & Tershy, B.R. (2015) Deep learning for large scale biodiversity monitoring. *Bloomberg Data for Good Exchange Conference*, pp. 1-7. Santa Cruz, USA.

Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conferences on Artificial Intelligence*, pp. 1137-1145. Montreal, Canada.

Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C. & Soares, J.V. (2012) Leafsnap: A computer vision system for automatic plant species identification. *European Conference on Computer Vision*, pp. 502-516. Florence, Italy.

MacLeod, N., Benfield, M. & Culverhouse, P. (2010) Time to automate identification. *Nature,* **467,** 154-155.

Molinaro, A.M., Simon, R. & Pfeiffer, R.M. (2005) Prediction error estimation: A comparison of resampling methods. *Bioinformatics,* **21,** 3301-3307.

Narins, P.M. & Capranica, R.R. (1977) An automated technique for analysis of temporal features in animal vocalizations. *Animal Behaviour,* **25,** 615-621.

Philippe, J., Felipe, L. & Celio, F. (2017) The use of bioacoustics in anuran taxonomy: Theory, terminology, methods and recommendations for best practice. *Zootaxa,* **4251,** 1-124.

Pili, A.N., Sy, E.Y., Diesmos, M.L.L. & Diesmos, A.C. (2019) Island hopping in a biodiversity hotspot archipelago: Reconstructed invasion history and updated status and distribution of alien frogs in the Philippines. *Pacific Science,* **73,** 321-343.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J. & Hughes, D.P. (2017) Deep learning for image-based cassava disease detection. *Frontiers in Plant Science,* **8,** 1852-1859.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A. & Bernstein, M. (2015) ImageNet large scale visual recognition challenge. *International Journal of Computer Vision,* **115,** 211-252.

Rzanny, M., Seeland, M., Wäldchen, J. & Mäder, P. (2017) Acquiring and preprocessing leaf images for automated plant identification: Understanding the tradeoff between effort and information gain. *Plant Methods,* **13,** 1-11.

Scheffers, B.R., Brunner, R.M., Ramirez, S.D., Shoo, L.P., Diesmos, A. & Williams, S.E. (2013) Thermal buffering of microhabitats is a critical factor mediating warming vulnerability of frogs in the Philippine biodiversity hotspot. *Biotropica,* **45,** 628-635.

Schmidhuber, J. (2015) Deep learning in neural networks: An overview. *Neural Networks,* **61,** 85-117.

Siler, C.D., Diesmos, A.C., Linkem, C.W., Diesmos, M.L. & Brown, R.M. (2010) A new species of limestone-forest frog, genus *Platymantis* (Amphibia: Anura: Ceratobatrachidae) from central Luzon Island, Philippines. *Zootaxa,* **2482,** 49-63.

Siler, C.D., Linkem, C.W., Diesmos, A.C. & Alcala, A.C. (2007) A new species of *Platymantis* (Amphibia: Anura: Ranidae) from Panay Island, Philippines. *Herpetologica*, **63,** 351-364.

Smith, S.W., Walsh, B., Grauer, K., Wang, K., Rapin, J., Li, J., Fennell, W. & Taboulet, P. (2019) A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *Journal of Electrocardiology,* **52,** 88-95.

Sueur, J., Aubin, T. & Simonis, C. (2008) Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics,* **18,** 213-226.

Sugai, L.S.M., Silva, T.S.F., Ribeiro Jr, J.W. & Llusia, D. (2019) Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience,* **69,** 15-25.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016) Rethinking the inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826. Las Vegas, USA.

Tapley, B., Michaels, C.J., Gumbs, R., Böhm, M., Luedtke, J., Pearce-Kelly, P. & Rowley, J.J. (2018) The disparity between species description and conservation assessment: A case study in taxa with high rates of species discovery. *Biological Conservation,* **220,** 209-214.

Taylor, E.H. (1923) Addition to the herpetological fauna of the Philippine Islands III. *Philippine Journal of Science,* **22,** 515-557.

Tonini, J.F.R., Provete, D.B., Maciel, N.M., Morais, A.R., Goutte, S., Toledo, L.F. & Pyron, R.A. (2020) Allometric escape from acoustic constraints is rare for frog calls. *Ecology and Evolution***,** 1-10.

Vieites, D.R., Wollenberg, K.C., Andreone, F., Köhler, J., Glaw, F. & Vences, M. (2009) Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of USA,* **106,** 8267-8272.

Vignal, C. & Kelley, D. (2007) Significance of temporal and spectral acoustic cues for sexual recognition in *Xenopus laevis*. *Proceedings of the Royal Society B,* **274,** 479-488.

Villa, A.G., Salazar, A. & Vargas, F. (2017) Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics,* **41,** 24-32.

Wells, K.D. (2010) *The ecology and behavior of amphibians*. University of Chicago Press.

Wells, K.D. & Schwartz, J.J. (2007) The behavioral ecology of anuran communication. *Hearing and sound communication in amphibians*, pp. 44-86. Springer, New York, NY.

Wimmer, J., Towsey, M., Roe, P. & Williamson, I. (2013) Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications,* **23,** 1419-1428.

Zhao, Z., Zhang, S., Xu, Z., Bellisario, K., Dai, N., Omrani, H. & Pijanowski, B.C. (2017) Automated bird acoustic event detection and robust species classification. *Ecological Informatics,* **39,** 99-108.

CHAPTER 3. Application of deep learning to citizen-science-based mosquito surveillance and detection of novel species

3.1 Introduction

Mosquitoes, which are responsible for 725,000 deaths per year globally (Gates 2014), are one of the most prominent insect disease vector groups (Juliano and Lounibos 2005). Two genera of particular concern to public health are *Aedes* and *Anopheles* (Lounibos and Kramer 2016, Tabbabi and Daaboub 2018), both because of the diseases they carry and because of their ability to travel and adapt to new environments. For instance, *Anopheles gambiae*, a vector of the malaria-causing parasite *Plasmodium* native to Africa, was discovered breeding in both Brazil and Greece in 1930 (Barber and Rice 1935, Shannon 1942). Similarly, *Aedes* species, vectors of zoonotic arbovirus diseases like yellow fever (Couto-Lima et al. 2017), Zika (Diagne et al. 2015), chikungunya (Vega-Rúa et al. 2014), and dengue (Lambrechts et al. 2010), have become established on several continents, leading to outbreaks of Zika in Brazil (Bogoch et al. 2016), chikungunya in Europe (Van Bortel et al. 2014), and dengue in Europe and the Americas (Adalja et al. 2012, Akiner et al. 2016).

As many people around the world are increasingly at risk of infection with these diseases, direct surveillance of mosquito populations is needed to provide appropriate and timely measures of containing and preventing the spread of mosquito-borne diseases. Studying mosquito wingbeats and their species- and gender-specific sounds is one of the most efficient ways of monitoring these populations. Though researchers have found success in focusing on acoustic approaches (Moore and Miller 2002, Li et al. 2005, Ouyang et al. 2015, Fanioudakis et al. 2018, Jansson et al. 2019), these methods are difficult to implement in the areas most affected by mosquito-borne diseases: a lack of financial resources often prohibits people in these areas from obtaining the expensive

equipment and expertise needed to monitor mosquito populations effectively. Therefore, methods that utilize inexpensive and accessible equipment, and that rely on the participation of citizen scientists, are much more likely to be successful surveillance systems.

Here, we explore a practical solution that uses smartphones for the acoustic monitoring of local mosquito species in Kansas, in the central United States. We focus specifically on smartphones because of their popularity and because of their utility: there are over three billion smartphone users worldwide (Statista; https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/), their built-in microphones are sensitive enough to be used for recording mosquito wingbeats, and their ability to collect recording metadata makes public health surveillance easy. We collected live mosquitoes, and recorded their wingbeats using smartphone microphones. Then, we applied a transfer-learning technique with Inception v3 (Szegedy et al. 2016), a robust, deep-learning-based model, to spectrograms (images representing simple acoustic features) generated from the wingbeat recordings to identify individuals at species level. Lastly, we challenged our model by introducing recordings of two potential invasive species, *Ae. aegypti* and *An. gambiae*, to our classification tasks.

3.2 Materials and Methods

3.2.1 Data collection and processing

During the 2018 and 2019 Kansas mosquito seasons (May to October), we collected recordings of 13 local mosquito species from 14 sites across Douglas County, Kansas. Seven of these sites were located in parkland areas, five in residential areas, one in a woodland area, and one in a wetland. In addition, we collected recordings of two potential invasive species, *Anopheles*

gambiae and *Aedes aegypti*, from a captive population in Accra, Ghana, and a disturbed forested area in Xalapa, Mexico, respectively.

We designed our mosquito recording methodology as practically as possible, such that any citizen scientist can successfully repeat the experiment. In doing so, we chose collecting and recording equipment that are inexpensive and easily accessible in any department store or from online suppliers. These items included thin-walled Ziploc bags, a standard fine mesh net for insects and butterflies, a small vial, dry ice to attract specific species, and a regular smartphone. For our project, smartphones included a Samsung Galaxy S6 SM-G920I and an iPhone SE model A1723.

To record mosquito wingbeats, we caught each live individual using the fine mesh net, taking precautions to avoid any physical damage to the mosquito's body. We used this method in place of other, more common methods (e.g., aspirators, vacuum-based devices) because those devices can damage the mosquito's wings or legs in ways that can introduce errors to our audio reference library. After catching an individual mosquito, we transferred it from the net to a Ziploc bag (where we eventually recorded its wingbeats) by first transferring it to a vial. To get the mosquito inside the vial, we gripped the base of the netting tightly, and held the vial against the net. Then, we slowly released the net, and flipped the vial and net together such that the vial came on top of the net and allowed the mosquito to fly into the vial. To transfer the mosquito from the vial to the Ziploc bag, we placed the vial into the Ziploc bag, holding the bag above the vial such that the mosquito could fly upward into the bag. Once the mosquito was inside the vial, we slowly removed the vial out and carefully zipped the bag to avoid any damage to the mosquito. We then carefully inflated the bag to provide enough space for the mosquito to fly and to avoid damaging the individual while transporting it from the collection site. To record their wingbeats, we took the mosquitoes to a quiet place (with no outside noise), and recorded using the default settings of each

smartphone's built-in microphone while mosquitos flew inside the inflated bags. We held the built-in microphone directly against the Ziploc bag, and recorded for 3-5 minutes. All recordings were generated in MPEG 4 Audio (M4A) format, an extension for audio files encoded with advance audio coding (AAC).

In total, we collected 358 individuals and 389 recordings, which resulted in ~1400 minutes of mosquito sounds. In addition to recordings, we catalogued secondary information associated with each mosquito. This information included collector's name, recording filename, date and time, type of attraction, country, city, location, habitat, number of recordings per individual, geographic coordinates, smartphone model, and species and gender (as identified by the entomologist member of our team).

We processed all recordings by first removing all recordings of male individuals: we focused specifically on females since male mosquitoes do not feed on blood and thus cannot transmit diseases; males have different wing beat characteristics (Simões et al. 2016, Mukundarajan et al. 2017, Lapshin and Vorontsov 2019), and were much less frequent in our collections than females. Then, we used the free, cross-platform audio editor, Ocenaudio (https://www.ocenaudio.com) to clip one-second segments of mosquito wingbeats. These segments were selected manually, and saved as 32-bit, single-channel WAV files (44.1 kHz sampling rate). We filtered out all segments that contained human noise, or that were silent (i.e., the mosquitos were not flying). In doing so, we retained as many high-quality, one-second wingbeat sounds as possible for each species.

We used R packages warbleR (Araya-Salas and Smith-Vidaurre 2017) and Seewave (Sueur et al. 2008) to generate spectrograms of the best-quality, one-second segments across a standardized range of frequencies, 0.1–1.1 kHz; all collected mosquito wingbeats fall within this

range (Figure 1). To generate spectrograms, we chose a fast-Fourier transformation (FFT) of 512 points, with 90% overlap between two successive windows. We saved all spectrograms as Portable Network Graphics (PNGs). We then visually checked all spectrograms to select 20 best-quality wingbeat spectrograms per species, i.e., we removed all spectrograms that indicated strong background or microphone noise. This reduced the number of species from 15 to 6: 4 local, and 2 invasive.

3.2.2 Classification tasks and model architecture

We designed four classification tasks to test performance of TensorFlow Inception v3 (Abadi et al. 2016, Szegedy et al. 2016) in identifying mosquito species, using recordings of their wingbeats. Inception v3 is a convolutional neural network (CNN) with 48 layers, implemented in TensorFlow (Abadi et al. 2016), and pre-trained on >1M images from the ImageNet database (Russakovsky et al. 2015). This model provides a transfer-learning technique, a shortcut for achieving high-performance classification using CNN models (Szegedy et al. 2016). In this way, we are able to re-train the final layer of Inception v3 on our training dataset, benefit from the experience already gained by this model, and avoid expensive computational processing to train a CNN from scratch (Qiu et al. 2017). To train the model, we modified two parameters: (1) number of training steps, and (2) validation percentage. We explored different numbers of training steps and compared results to find an optimum balance between computing time and classification efficiency. Given our limited number of images per species (n=20), we used a validation parameter of 20% in all classification tasks except for Task 4, in which we chose 10%. In all tasks, we used a leave-one-out cross validation technique (Molinaro et al. 2005) to evaluate model performance, as in our previous work (Khalighifar et al. 2019, Khalighifar et al. 2020b).

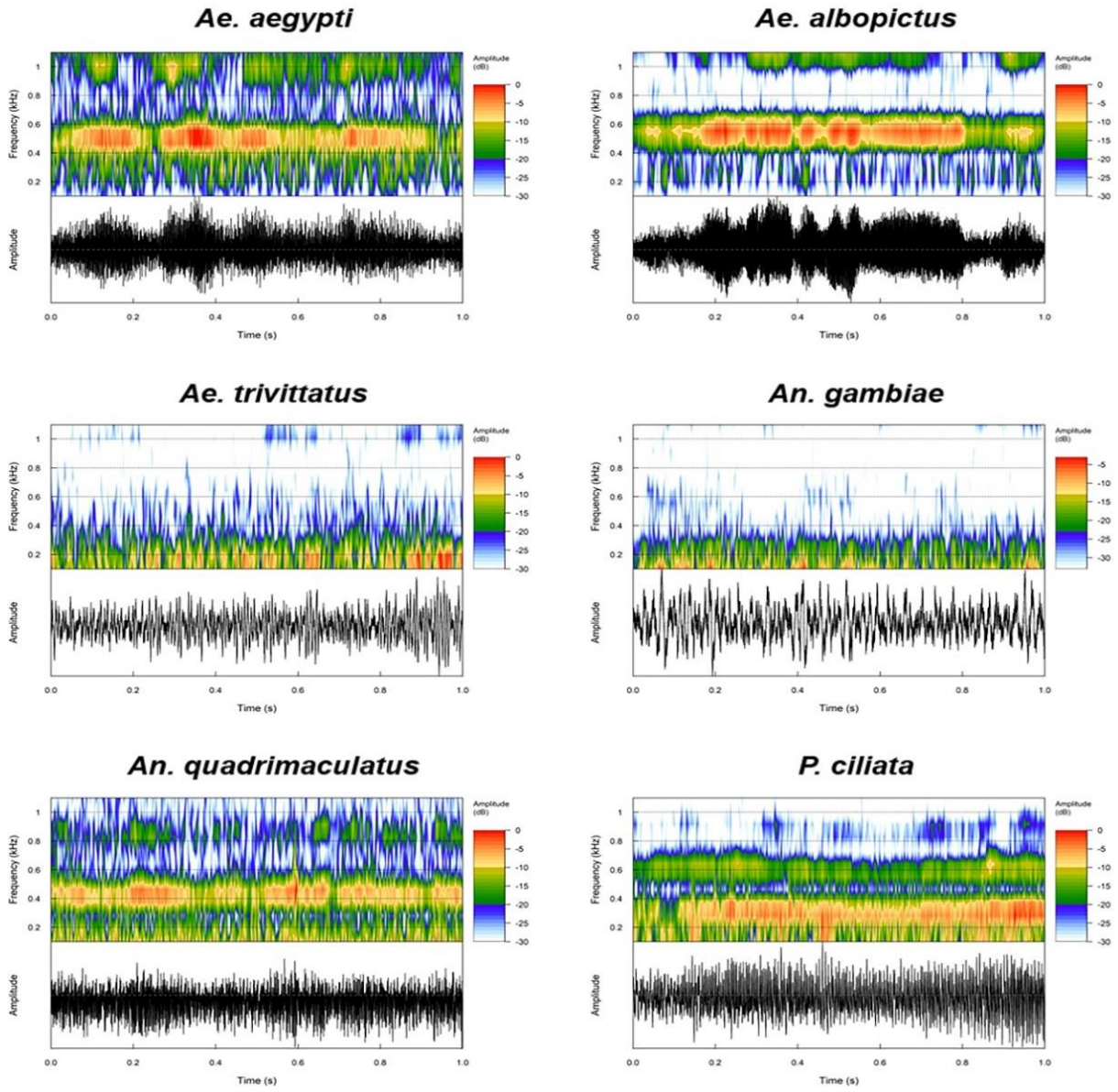

Figure 1. Examples of spectrogram images used for identification of mosquito species based on their wingbeats recorded by smartphones. Each spectrogram has the same time duration (one second), and frequency limits (0.1-1.1 kHz).

The four classification tasks we explored are as follows:

Task 1. Applying TensorFlow to identify local mosquito species in Kansas: We applied TensorFlow to wingbeat recordings from four of the most common local species in Kansas to test the model's performance in species identification. We selected the best 20 spectrograms per species (80 in total) for this classification task.

Task 2. Challenging TensorFlow with recordings associated with species not present in the reference library: We trained TensorFlow on all spectrograms associated with local species (i.e., the 80 images in the reference library from Task 1). Then, we applied the model to a test dataset containing spectrograms of our two potential invasive species, *An. gambiae* and *Ae. aegypti*, to address how TensorFlow performs when it encounters species that do not exist in the reference library. To set up this task, we selected the best 20 spectrograms for each of the potential invasive species, resulting in test data set containing 40 images.

Task 3. Applying TensorFlow to wingbeat recordings of all six species: We added recordings associated with two potential invasive species (i.e., those form Task 2) to the reference library of Task 1, to test TensorFlow's performance in distinguishing the potential invasive species from the local species. As such, the reference library for this classification task contained 120 images (again, 20 spectrograms per species).

Task 4. Applying TensorFlow to recording collected by different smartphones: As we used two different smartphone brands/models to record mosquito sounds, we used this final classification task to address (1) if TensorFlow is able to differentiate recordings collected by different phones, and (2) if so, whether smartphone model impacts the accuracy of our species identification process. Here, we created an image reference library consisting of 100 randomly-selected spectrograms per smartphone model (200 images in total).

3.3 Results

To develop a robust and accessible mosquito identification system, we generated smartphone recordings of 13 local and 2 potentially invasive species. We applied two rounds of cleaning to our dataset: first, manually clipping the best-quality, one-second segments of female mosquito sounds, and then manually selecting the 20 best-quality spectrogram images per species. This resulted in a total of 6 mosquito species, 4 local and 2 potential invasive, that had sample sizes sufficient to be used in four classification tasks designed to challenge TensorFlow's ability to identify mosquitoes at species level using wingbeat sounds. The details of results associated with each classification task are as follows:

Species Name	*Ae. albopictus*	*Ae. trivittatus*	*An. quadrimaculatus*	*P. ciliata*
Ae. albopictus	**0.75**		0.1	0.15
Ae. trivittatus		**0.95**	0.05	
An. quadrimaculatus	0.1		**0.8**	0.1
P. ciliata		0.05		**0.95**

Figure 2. Confusion matrix for four local species in Douglas County, Kansas, using a leave-one-out cross-validation technique. Green colors represent correct identifications, and orange misidentifications. All values of zero are removed for ease of visualization.

Task 1. We applied TensorFlow to recordings associated with four local species (20 spectrograms per species), setting 2500 for training steps and 20% for validation. The model calibration for this and all following tasks were based on trials using different numbers of training steps, considering both correct identification rates and processing times to find optima for training TensorFlow (Khalighifar et al. 2019). We created a confusion matrix to depict TensorFlow's

performance in this task (Figure 2). The overall correct identification rate was 86.3%, and two species, *Aedes trivittatus* and *Psorophora ciliata*, were identified at a rate of 95.0%. *Aedes albopictus*, with a rate of 75.0%, had the lowest correct identification rate.

Task 2. TensorFlow provides two pieces of information associated with each image: suggested species names and certainty rate. Certainty rate can be a factor with which to evaluate classifier performance on test images, particularly as regards novel species not represented in the reference library (Khalighifar et al. 2020b). We trained TensorFlow on images associated with local species, once again using 2500 training steps and 20% validation, and then applied it to test images associated with two potential invasive species, *Ae. aegypti* and *An. gambiae*. Since these two species were not present in the reference library, it was impossible for TensorFlow to identify them correctly. However, TensorFlow was consistent in classifying test images: the model identified all *Ae. aegypti*'s images as *Ae. albopictus*, and all *An. gambiae*'s images as *Ae. trivittatus* (Figure 3). The overall certainty rate for these two species was 93.2% (95.6% for *Ae. aegypti* and 90.7% for *An. gambiae*), meaning that TensorFlow showed a high confidence in misidentifying the two novel species.

Species Names	*Ae. albopictus*	*Ae. trivittatus*	*An. quadrimaculatus*	*P. ciliata*	Certainty Rate
Ae. aegypti	20 95.6%				95.6%
An. gambiae		20 90.7%			90.7%

Figure 3. Confusion matrix resulting from challenging TensorFlow with potential invasive species unknown to the reference library. Columns represent species in the reference library, rows are potential invasive species. Red numbers show the high certainty rates generated by TensorFlow. The far-right column is the average certainty rate for species identifications.

Based on previous experience with frog calls (Khalighifar et al. 2020b), we would expect lower certainty rates in Task 2: species not in the reference library should have much lower certainty rates than those already present in the library. We did not see this pattern. However, we did see this kind of pattern with misidentifications in Task 1, where the overall certainty rate for misidentifications, 61.8%, was significantly lower than the certainty rate of correctly identified images, 88.8% (Mann-Whitney U=78.0, P=1.30e-05; Python 3.8.2). In Task 1, we achieved overall certainty rates of 90.9% and 89.1% for correctly identified images associated with *Ae. trivittatus* and *Ae. albopictus*, respectively. These numbers are approximately close to certainty rates associated with *An. gambiae* (90.7%) and *Ae. aegypti* (95.6%) in this task, which were misidentified as *Ae. trivittatus* and *Ae. albopictus*, respectively.

Task 3. Here, we added images associated with *Ae. aegypti* and *An. gambiae* to the reference library to test TensorFlow's performance ability to discriminate two potentially invasive species against local species in a model with 4000 training steps and 20% validation. We created a confusion matrix to show correct identification rates associated with each species (Figure 4). The overall correct identification rate was 85.0%. We achieved correct identification of 95% or above for three species. Most notable among these three, *An. gambiae* was identified with 100% correct identification rate. *Ae. albopictus* and *Ae. aegypti* were the most challenging species to TensorFlow, and showed the lowest identification rates: 70% of images (14 out of 20) were identified correctly. All six misidentified images associated with *Ae. aegypti* were classified as *Ae. albopictus*.

Task 4. We created an image reference library consisting of 100 randomly-selected spectrograms per smartphone model (200 images in total). We chose 1500 training steps and 10% validation to test TensorFlow's performance in distinguishing recordings collected by different

phones. The overall correct identification rate was 94.5%: images associated with the Samsung model had a 97% correct identification rate, and images associated with the iPhone model had a 92% correct identification rate.

Species Name	*Ae. aegypti*	*Ae. albopictus*	*Ae. trivittatus*	*An. gambiae*	*An. quadrimaculatus*	*P. ciliata*
Ae. aegypti	**0.7**	0.3				
Ae. albopictus	0.1	**0.7**			0.1	0.1
Ae. trivittatus			**0.95**		0.05	
An. gambiae				**1**		
An. quadrimaculatus		0.1			**0.8**	0.1
P. ciliata			0.05			**0.95**

Figure 4. Confusion matrix for all species combined (4 local species and 2 potential invasive) using a leave-one-out cross-validation technique. Green colors represent correct identifications, and orange misidentifications. All values of zero are removed for ease of visualization

Task 4. We created an image reference library consisting of 100 randomly-selected spectrograms per smartphone model (200 images in total). We chose 1500 training steps and 10% validation to test TensorFlow's performance in distinguishing recordings collected by different phones. The overall correct identification rate was 94.5%: images associated with the Samsung model had a 97% correct identification rate, and images associated with the iPhone model had a 92% correct identification rate.

To determine if smartphone model introduced bias to our analyses, we also measured the percent of images per species collected by each smartphone (Table 1). All recordings associated with *An. gambiae* and *Ae. aegypti* were collected using the iPhone. Although all images associated

with *An. gambiae* were identified as *Ae. trivittatus* in Task 2, only 40% of *Ae. trivittatus* recordings were collected by the iPhone. Even though Task 2 and 3 showed that *Ae. aegypti* and *Ae. albopictus* were challenging to TensorFlow, 25% of *Ae. albopictus* recordings were collected using the Samsung. Additionally, among the six misidentifications associated with *Ae. albopictus* class, three recordings were collected with the iPhone. We collected 85% of *P. ciliata* recordings with the Samsung and the remaining 15% with the iPhone. Despite the majority of *P. ciliata* recordings being taken with the same phone model, the only misidentification associated with this species was collected with the Samsung. To this end, although TensorFlow was able to distinguish spectrograms associated with different smartphone models, we believe differences between models did not introduce any bias to our analyses because no systematic misidentifications related to either model was detected.

Table 1. Number of spectrograms obtained from recordings collected by different smartphones per species.

	Total number of images per smartphone		**Total number of misidentifications per smartphone**	
Species name	**iPhone**	**Samsung**	**iPhone**	**Samsung**
Ae. aegypti	20	-	6	-
Ae. albopictus	15	5	3	3
Ae. trivittatus	8	12	-	1
An. gambiae	20	-	-	-
An. quadrimaculatus	-	20	-	4
P. ciliata	3	17	-	1

3.4 Discussion

In this project, we proposed a practical and inexpensive, citizen-science-based approach for improving surveillance of mosquito populations. We used smartphones to record mosquito

wingbeats, and then applied state-of-the-art deep-learning techniques to identify recordings at species level. In this proof-of-concept exploration, we show that only 20 spectrograms per species can achieve an 85.0% overall correct identification rate for six species, with three of those species being correctly identified at a rate of 95.0% or above. Though our sample size is relatively small, these identification rates are markedly higher than those achieved by previous studies using expensive devices to record wingbeats (Moore and Miller 2002, Li et al. 2005, Ouyang et al. 2015, Jansson et al. 2019). Though adding more species to our reference library and increasing per-species sample size would allow us to utilize more robust model evaluation techniques (i.e., *k*-fold cross-validation) and increase the overall accuracy of our model, we believe that these results demonstrate that our approach has promise as a citizen-science-implemented mosquito surveillance system.

However, though TensorFlow was able to give the 'best possible guess' for *Ae. aegypti* and *An. gambiae* in Task 2, the high certainty rates associated with these identifications indicates that TensorFlow was unable to flag recordings associated with novel species. The fact that these novel species were not present in the reference library means that it was impossible for TensorFlow to identify them correctly. However, results of our previous work led to us to expect significantly lower certainty rates for images associated with species absent from our reference library. In our previous studies, TensorFlow was able to flag species new to the reference library (Khalighifar et al. 2020b) and mislabeled species already present in the reference library (Khalighifar et al. 2020a) with lower certainty rates for images associated with those species. Though *Ae. aegypti* and *Ae. albopictus* spectrograms, as well as *An. gambiae* and *Ae. trivittatus* spectrograms, are very similar to each other (Figure 1), we believe that significantly increasing the sample size of each of these

species would introduce more information about wingbeat pattern that would allow TensorFlow to differentiate between them.

In Task 4, TensorFlow was able to differentiate between Samsung and iPhone recordings with a 94.5% overall correct identification rate. We believe TensorFlow was able to distinguish between Samsung and iPhone recordings because of systematic noises generated by the Samsung; although many smartphones now contain advanced, built-in microphones, they cannot generate the same, high-quality recordings produced with high-end recording equipment. Differences between smartphones, however, did not appear to introduce any bias to the identification process. This finding is echoed in Mukundarajan et al. (2017), who also showed that differences between cell phone models do not impact the identification process. As such, we believe that adding wingbeat recordings from multiple smartphone models to our reference library would add the smartphone recording variation necessary to an accurate, future citizen-science-based identification system.

In conclusion, we believe our proof-of-concept shows high potential for the combination of robust deep-learning techniques and smartphone recording technology to monitor mosquito populations. Our method is practical and easy-to-learn, such that any citizen scientist with access to a smartphone and inexpensive collecting equipment would be able to contribute to this kind of surveillance system. Significant challenges exist as regards detection of novel species, and their characterization as novel, but we are optimistic that these problems may be resolved with much-increased sample sizes. As climate change continues to alter the distribution of mosquito populations throughout the Americas (Reiter 2001), our work provides a simple and efficient monitoring method for offsetting the enormous public health costs of mosquito-borne diseases (Suaya et al. 2007).

3.5 References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard. 2016. Tensorflow: A system for large-scale machine learning, pp. 265-283, 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, USA.

Adalja, A. A., T. K. Sell, N. Bouri, and C. Franco. 2012. Lessons learned during dengue outbreaks in the United States, 2001–2011. Emerging Infectious Diseases 18: 608-614.

Akiner, M. M., B. Demirci, G. Babuadze, V. Robert, and F. Schaffner. 2016. Spread of the invasive mosquitoes *Aedes aegypti* and *Aedes albopictus* in the Black Sea region increases risk of chikungunya, dengue, and Zika outbreaks in Europe. PLoS Neglected Tropical Diseases 10: e0004664.

Araya-Salas, M., and G. Smith-Vidaurre. 2017. warbleR: An R package to streamline analysis of animal acoustic signals. Methods in Ecology and Evolution 8: 184-191.

Barber, M., and J. Rice. 1935. Malaria studies in Greece: The malaria infection rate in nature and in the laboratory of certain species of *Anopheles* of East Macedonia. Annals of Tropical Medicine & Parasitology 29: 329-348.

Bogoch, I. I., O. J. Brady, M. U. Kraemer, M. German, M. I. Creatore, M. A. Kulkarni, J. S. Brownstein, S. R. Mekaru, S. I. Hay, and E. Groot. 2016. Anticipating the international spread of Zika virus from Brazil. Lancet 387: 335-336.

Couto-Lima, D., Y. Madec, M. I. Bersot, S. S. Campos, M. de Albuquerque Motta, F. B. Dos Santos, M. Vazeille, P. F. da Costa Vasconcelos, R. Lourenço-de-Oliveira, and A.-B. Failloux. 2017. Potential risk of re-emergence of urban transmission of yellow fever virus in Brazil facilitated by competent *Aedes* populations. Scientific Reports 7: 4848.

Diagne, C. T., D. Diallo, O. Faye, Y. Ba, O. Faye, A. Gaye, I. Dia, S. C. Weaver, and M. Diallo. 2015. Potential of selected Senegalese *Aedes* spp. mosquitoes (Diptera: Culicidae) to transmit Zika virus. BMC Infectious Diseases 15: 492

Fanioudakis, E., M. Geismar, and I. Potamitis. 2018. Mosquito wingbeat analysis and classification using deep learning, pp. 2410-2414, 26th European Signal Processing Conference. IEEE, Rome, Italy

Gates, B. 2014. The deadliest animal in the world (https://www.gatesnotes.com/Health/Most-Lethal-Animal-Mosquito-Week; accessed in July, 2020).

Jansson, S., A. Gebru, R. Ignell, J. Abbott, and M. Brydegaard. 2019. Correlation of mosquito wing-beat harmonics to aid in species classification and flight heading assessment, pp. 11075_11024, European Conference on Biomedical Optics. Optical Society of America, Munich, Germany.

Juliano, S. A., and L. P. Lounibos. 2005. Ecology of invasive mosquitoes: Effects on resident species and on human health. Ecology Letters 8: 558-574.

Khalighifar, A., S. E. Little, and A. T. Peterson. 2020a. Accelerating surveillance of tick-borne diseases via deep learning technology. Journal of Medical Entomology: *(In preparation)*.

Khalighifar, A., R. Brown, J. Goyes Vallejos, and A. T. Peterson. 2020b. Deep learning technology improves auditory biodiversity assessment and new candidate species identification in a hyper-diverse and yet underestimated species assemblage from an island archipelago. Methods in Ecology and Evolution: *(In review)*.

Khalighifar, A., E. Komp, J. M. Ramsey, R. Gurgel-Gonçalves, and A. T. Peterson. 2019. Deep learning algorithms improve automated identification of Chagas disease vectors. Journal of Medical Entomology 56: 1404-1410.

Lambrechts, L., T. W. Scott, and D. J. Gubler. 2010. Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. PLoS Neglected Tropical Diseases 4: e646.

Lapshin, D. N., and D. D. Vorontsov. 2019. Directional and frequency characteristics of auditory neurons in *Culex* male mosquitoes. Journal of Experimental Biology 222: jeb208785.

Li, Z., Z. Zhou, Z. Shen, and Q. Yao. 2005. Automated identification of mosquito (Diptera: Culicidae) wingbeat waveform by artificial neural network, pp. 483-489, International Conference on Artificial Intelligence Applications and Innovations. Springer, Beijing, China.

Lounibos, L. P., and L. D. Kramer. 2016. Invasiveness of *Aedes aegypti* and *Aedes albopictus* and vectorial capacity for chikungunya virus. Journal of Infectious Diseases 214: 453-458.

Molinaro, A. M., R. Simon, and R. M. Pfeiffer. 2005. Prediction error estimation: A comparison of resampling methods. Bioinformatics 21: 3301-3307.

Moore, A., and R. H. Miller. 2002. Automated identification of optically sensed aphid (Homoptera: Aphidae) wingbeat waveforms. Annals of the Entomological Society of America 95: 1-8.

Mukundarajan, H., F. J. H. Hol, E. A. Castillo, C. Newby, and M. Prakash. 2017. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. eLife 6: e27854.

Ouyang, T.-H., E.-C. Yang, J.-A. Jiang, and T.-T. Lin. 2015. Mosquito vector monitoring system based on optical wingbeat classification. Computers and Electronics in Agriculture 118: 47-55.

Qiu, Z., T. Yao, and T. Mei. 2017. Learning spatio-temporal representation with pseudo-3d residual networks, pp. 5533-5541, IEEE International Conference on Computer Vision, Venice, Italy.

Reiter, P. 2001. Climate change and mosquito-borne diseases. Environmental Health Perspectives 109: 141-161.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. 2015. ImageNet large scale visual recognition challenge. International Journal of Computer Vision 115: 211-252.

Shannon, R. C. 1942. Brief history of *Anopheles gambiae* in Brazil. Caribbean Medical Journal 4: 123-128.

Simões, P. M., R. A. Ingham, G. Gibson, and I. J. Russell. 2016. A role for acoustic distortion in novel rapid frequency modulation behaviour in free-flying male mosquitoes. Journal of Experimental Biology 219: 2039-2047.

Suaya, J. A., D. S. Shepard, and M. E. Beatty. 2007. Dengue: Burden of disease and costs of illness, pp. 35-49, Scientific Working Group on Dengue Research, Geneva, Switzerland.

Sueur, J., T. Aubin, and C. Simonis. 2008. Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics 18: 213-226.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. Rethinking the inception architecture for computer vision, pp. 2818-2826, IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA.

Tabbabi, A., and J. Daaboub. 2018. Fitness cost in field *Anopheles* labranchiae populations associated with resistance to the insecticide deltamethrin. Revista Brasileira de Entomologia 62: 107-111.

Van Bortel, W., F. Dorleans, J. Rosine, A. Blateau, D. Rousset, S. Matheus, I. Leparc-Goffart, O. Flusin, C. Prat, and R. Cesaire. 2014. Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. Eurosurveillance 19: 20759.

Vega-Rúa, A., K. Zouache, R. Girod, A.-B. Failloux, and R. Lourenço-de-Oliveira. 2014. High level of vector competence of *Aedes aegypti* and *Aedes albopictus* from ten American countries as a crucial factor in the spread of chikungunya virus. Journal of Virology 88: 6294-6306.

CONCLUSION AND FUTURE DIRECTIONS

The set of comprehensive analyses presented here represents a new application of deep-learning techniques to investigations of biological research questions. In developing these models, we provide researchers with examples and illustrations of the potential utility of a high-end technological tool that can be used to address questions on multiple aspects of biodiversity and ecology. While these research questions include a variety of topics under computational biology, biodiversity science, ecology, and biogeography, our models focus most specifically around rigorous species identification and detecting novel biodiversity. In this "conclusions" chapter, I explore at least briefly a set of future directions and research challenges to which these novel tools can be applied.

Although research focus on cryptic species has grown impressively over the past two decades, such species remain one of the most significant challenges in systematic biology (Bickford et al. 2007). Taxonomic impediments do not allow easy testing of hypotheses about crypticity. For instance, under current taxonomic methods, it is difficult to test whether cryptic species are more common in particular habitats, latitudes, or taxonomic groups. As demonstrated in Chapters 1 and 2, automated species identification platforms can provide a robust foundation for addressing these questions, given the CNNs' ability to use overall signal to identify species. For example, in addition to discriminating the Chagas disease vectors with the highest prevalence in Mexico and Brazil at the species level, TensorFlow was able to discriminate species belonging to the *Triatoma dimidiata* complex, a set of cryptic lineages previously impossible to identify morphologically (Cruz et al. 2020) and generally identified only via DNA barcoding approaches. In our work, these three lineages were identified with a correct overall identification rate of ~86%. Another example is with vocally cryptic species, frogs—until recently, three species in our

analyses (*Platymantis banahao*, *P. luzonensis*, and *P. rabori*) were considered to be part of *P. guentheri* (Brown et al. 2015b). TensorFlow, however, was able to discriminate all four of these taxa with a correct overall identification rate of 96.3%, as well as differentiate between species with similar call properties and close, pairwise phylogenetic relationships with >95% overall correct identification rate (e.g., *P. indeprensus* and *P. mimulus*, *P. levigatus* and *P. insulatus*). In these examples, the high accuracy of CNNs in differentiating cryptic species makes automated species identification systems a powerful tool to answer questions about morphologically and genetically similar species, and to address more broad questions regarding crypticity.

Detecting undescribed species or the presence of novel, invasive species represent other challenges for researchers. As with cryptic species, taxonomic impediments and the need for experts to assess images, recordings, and specimens manually dramatically slows identification of new species. Similar obstacles exist for monitoring invasive species: researchers must survey for species of interest, and then track their spread manually. As explored in Chapters 2 and 3, automated identification systems also have the potential to increase the speed and accuracy of new species discovery and invasive species detection. For instance, TensorFlow was able to detect frog mate calls new to the reference library by flagging novel spectrograms with a markedly lower certainty rate, and was able to distinguish *An. gambiae* and *Ae. aegypti* from all local mosquito species with 85% overall correct identification rate (although the certainty rates were not lower as in the *Platymantis* example). A greater ability to track novel species would allow researchers to ask more informed questions about the ecology of places where significant portions of the biodiversity are unknown, and allow them to learn more about invasive species' distributions that could help mitigate the economic and ecological costs of establishment.

Additional future steps for full development of these platforms that I plan to explore can be distilled down to four general types of research questions. The first deals with practical applications of these models, such as long-term, large-scale monitoring of target species and monitoring individuals in species' populations based on specific characteristics that can be observed in each individual; for example, color patterns in *Espadarana prosoblepon* (Basto-Riascos et al. 2017), and unique stripe patterns in *Equus grevyi* (Zero et al. 2013). With a large-scale, monitoring cyberinfrastructure in place, it is also possible to test synthetic hypotheses regarding species' range shifts in the face of global change (Williams and Blois 2018, Mortelliti et al. 2019), or biotic homogenization and associated reductions in beta-diversity as a consequence of ecosystem disturbance (Beauvais et al. 2016, Price et al. 2020). Deep-learning models could also be used to answer behavioral biology questions. For instance, these models can help to understand better the signals used by species showing Batesian mimicry (e.g., *Limenitis archippus* mimicking *Danaus plexippus*; Ritland and Brower 2000) or needing to recognize relatives as part of systems purportedly driven by kin selection (e.g., *Aphelocoma ultramarine*; Brown 1972). Lastly, biological research questions could explore how TensorFlow and other CNN models are *able to* perceive the patterns distinguishing cryptic forms, and as a consequence illuminate the signals that do exist and may be used by the individuals of those populations for recognition.

In summary, automated species identification systems like those I have explored in this dissertation offer an efficient and highly accurate method to improve current biodiversity monitoring routines. In allowing researchers to ask more specific questions about complex ecological processes and in encouraging participation of citizen scientists, these platforms not only further broad-scale biodiversity research endeavors, but also may be a vehicle by which to bring biodiversity science to members of the broader community. As such, these systems have high

potential to revolutionize ways that scientists are able to approach grand challenge questions in biodiversity conservation, ecology, evolution, and public health.

REFERENCES

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard. 2016. Tensorflow: A system for large-scale machine learning, pp. 265-283, 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, USA.

Ascensão, F., C. Branquinho, and E. Revilla. 2020. Cars as a tool for monitoring and protecting biodiversity. Nature Electronics 3: 295-297.

Barata, I. M., R. A. Griffiths, and M. S. Ridout. 2017. The power of monitoring: Optimizing survey designs to detect occupancy changes in a rare amphibian population. Scientific Reports 7: 16491.

Basto-Riascos, M. C., J. López-Caro, and F. Vargas-Salinas. 2017. Reproductive ecology of the glass frog *Espadarana prosoblepon* (Anura: Centrolenidae) in an urban forest of the Central Andes of Colombia. Journal of Natural History 51: 2535-2550.

Beauvais, M. P., S. Pellerin, and C. Lavoie. 2016. Beta diversity declines while native plant species richness triples over 35 years in a suburban protected area. Biological Conservation 195: 73-81.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148-155.

Blumstein, D. T., D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, and K. A. Cortopassi. 2011. Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations, and prospectus. Journal of Applied Ecology 48: 758-767.

Bogich, T. L., A. M. Liebhold, and K. Shea. 2008. To sample or eradicate? A cost minimization model for monitoring and managing an invasive species. Journal of Applied Ecology 45: 1134-1142.

Brown, J. L. 1972. Communal feeding of nestlings in the Mexican jay (*Aphelocoma ultramarina*): Interflock comparisons. Animal Behaviour 20: 395-403.

Brown, R., C. Dolino, E. Alcala, A. Diesmos, and A. Alcala. 2002. The advertisement calls of two endangered species of endemic Philippine frogs: *Platymantis spelaeus* and *P. insulatus* (Anura; Ranidae). Silliman Journal 43: 91-109.

Brown, R. M., and J. C. Gonzalez. 2007. A new forest frog of the genus *Platymantis* (Amphibia: Anura: Ranidae) from the Bicol Peninsula of Luzon Island, Philippines. Copeia 2007: 251-266.

Brown, R. M., L. A. De Layola, A. Lorenzo, M. L. L. Diesmos, and A. C. Diesmos. 2015a. A new species of limestone karst inhabiting forest frog, genus *Platymantis* (Amphibia: Anura: Ceratobatrachidae: subgenus *Lupacolus*) from southern Luzon Island, Philippines. Zootaxa 4048: 191-210.

Brown, R. M., C. D. Siler, S. J. Richards, A. C. Diesmos, and D. C. Cannatella. 2015b. Multilocus phylogeny and a new classification for Southeast Asian and Melanesian forest frogs (family Ceratobatrachidae). Zoological Journal of the Linnean Society 174: 130-168.

Brown, R. M., A. C. Diesmos, M. B. Sanguila, C. D. Siler, M. L. Diesmos, and A. C. Alcala. 2012. Amphibian conservation in the Philippines. FrogLog 20: 40-43.

Carvalho, G., S. Creer, M. J. Allen, F. Costa, C. Tsigenopoulos, M. Le Goff-Vitry, A. Magoulas, L. Medlin, and K. Metfies. 2010. Genomics in the discovery and monitoring of marine biodiversity, pp. 1-32. In K. Tessmar-Raible, C. Boyen and F. Viard (eds.), Introduction to marine genomics, vol. 1. Springer, Dordrecht, Netherlands.

Chen, T., M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang. 2015. MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems. Distributed, Parallel, and Cluster Computing arXiv: 1512.01274.

Cruz, D. D., E. Arellano, D. D. Ávila, and C. N. Ibarra-Cerdeña. 2020. Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: A case for the cryptic *dimidiata* complex. Parasites and Vectors 13: 332.

Diesmos, A. C., J. L. Watters, N. A. Huron, D. R. Davis, A. C. Alcala, R. I. Crombie, L. E. Afuang, G. Gee-Das, R. V. Sison, and M. B. Sanguila. 2015. Amphibians of the Philippines, part I: Checklist of the species. Proceedings of the California Academy of Sciences 62: 457-539.

Hill, A. P., P. Prince, E. Piña Covarrubias, C. P. Doncaster, J. L. Snaddon, and A. Rogers. 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods in Ecology and Evolution 9: 1199-1211.

Ketkar, N. 2017. Introduction to pytorch, pp. 195-208. In T. Arritola, J. Gaines, A. Dragosavljević and T. Taylor (eds.), Deep learning with python. Apress, Berkeley, USA.

Klein, D. J., M. W. McKown, and B. R. Tershy. 2015. Deep learning for large scale biodiversity monitoring, pp. 1-7, Bloomberg Data for Good Exchange Conference, New York City, USA.

Lundervold, A. S., and A. Lundervold. 2019. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik 29: 102-127.

Mortelliti, A., I. P. Grentzmann, S. Fraver, A. M. Brehm, S. Calkins, and N. Fisichelli. 2019. Small mammal controls on the climate-driven range shift of woody plant species. Oikos 12: 1726-1738.

Mukundarajan, H., F. J. H. Hol, E. A. Castillo, C. Newby, and M. Prakash. 2017. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. eLife 6: e27854.

Nugent, J. 2018. iNaturalist: Citizen science for 21st-century naturalists. Science Scope 41: 12-13.

Ogden, E. L. 2018. Sounds good? The acoustic monitoring of coral-reef health. BioScience 68: 48-49.

Papeş, M., R. Tupayachi, P. Martinez, A. Peterson, and G. Powell. 2010. Using hyperspectral satellite imagery for regional inventories: A test with tropical emergent trees in the Amazon Basin. Journal of Vegetation Science 21: 342-354.

Price, E. P. F., G. Spyreas, and J. W. Matthews. 2020. Biotic homogenization of wetland vegetation in the conterminous United States driven by *Phalaris arundinacea* and anthropogenic disturbance. Landscape Ecology 35: 779-792.

Rassi, A., Jr., A. Rassi, and J. A. Marin-Neto. 2010. Chagas disease. Lancet 375: 1388-1402.

Reiter, P. 2001. Climate change and mosquito-borne diseases. Environmental Health Perspectives 109: 141-161.

Ritland, D. B., and L. P. Brower. 2000. Mimicry-related variation in wing color of viceroy butterflies (*Limenitis archippus*): A test of the model-switching hypothesis (Lepidoptera: Nymphalidae). Holarctic Lepidoptera 7: 42866.

Rowcliffe, M. J., C. Carbone, P. A. Jansen, R. Kays, and B. Kranstauber. 2011. Quantifying the sensitivity of camera traps: An adapted distance sampling approach. Methods in Ecology and Evolution 2: 464-476.

Rowley, J., R. Brown, R. Bain, M. Kusrini, R. Inger, B. Stuart, G. Wogan, N. Thy, T. Chan-ard, C. T. Trung, A. C. Diesmos, D. T. Iskandar, M. Lau, S. Makchai, N. Q. Truong, and S. Phimmachak. 2009. Impending conservation crisis for Southeast Asian amphibians. Biology Letters 6: 336-338.

Schmeller, D. S., M. Böhm, C. Arvanitidis, S. Barber-Meyer, N. Brummitt, M. Chandler, E. Chatzinikolaou, M. J. Costello, H. Ding, and J. García-Moreno. 2017. Building capacity in biodiversity monitoring at the global scale. Biodiversity and Conservation 26: 2765-2790.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. Neural Networks 61: 85-117.

Smith, S. W., B. Walsh, K. Grauer, K. Wang, J. Rapin, J. Li, W. Fennell, and P. Taboulet. 2019. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. Journal of Electrocardiology 52: 88-95.

Smolinski, M. S., A. W. Crawley, J. M. Olsen, T. Jayaraman, and M. Libel. 2017. Participatory disease surveillance: Engaging communities directly in reporting, monitoring, and responding to health threats. JMIR Public Health and Surveillance 3: e62.

Villanueva-Rivera, L. J., and B. C. Pijanowski. 2012. Pumilio: A web-based management system for ecological recordings. The Bulletin of the Ecological Society of America 93: 71-81.

Wang, R., and J. A. Gamon. 2019. Remote sensing of terrestrial plant biodiversity. Remote Sensing of Environment 231: 111218.

Webster, M. S., and G. F. Budney. 2017. Sound archives and media specimens in the 21st century, pp. 462-485. In C. Brown and T. Riede (eds.), Comparative bioacoustics: An overview. Bentham Science, Sharjah, UAE.

Williams, J. E., and J. L. Blois. 2018. Range shifts in response to past and future climate change: Can climate velocities and species' dispersal capabilities explain variation in mammalian range shifts? Journal of Biogeography 45: 2175-2189.

Yamashita, R., M. Nishio, R. K. G. Do, and K. Togashi. 2018. Convolutional neural networks: An overview and application in radiology. Insights into Imaging 9: 611-629.

Zero, V. H., S. R. Sundaresan, T. G. O'Brien, and M. F. Kinnaird. 2013. Monitoring an endangered savannah ungulate, Grevy's zebra *Equus grevyi*: Choosing a method for estimating population densities. Oryx 47: 410-419.

Zhang, J., J. Hu, J. Lian, Z. Fan, X. Ouyang, and W. Ye. 2016. Seeing the forest from drones: Testing the potential of lightweight drones as a tool for long-term forest monitoring. Biological Conservation 198: 60-69.