

DETICKT IT: An iOS App for Real-Time Tick Classification and Disease Risk Analysis

Antonia M. Kolb, Victoria K. Schulman
King School, 1450 Newfield Ave, Stamford, CT 06905

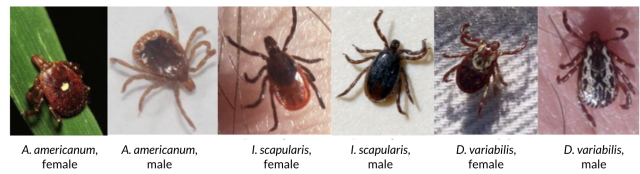
There is an alarming increase in the population of ticks and tick-borne diseases (TBDs), some of which are fatal. Due to limited training, healthcare providers cannot always accurately identify ticks and their associated illnesses, leading to delayed diagnoses and treatments. The prevalence rates of different disease-causing pathogens vary based on geographic locations within the United States. A convolutional neural network (CNN) was built for combining real-time tick-species identification with location-based tick-risk assessment by embedding the Pennsylvania Tick Research Lab's spatio-temporal tick surveillance statistics. With DETICKT IT, an iOS app developed in Swift with a Python backend, users can now receive an accurate and conclusive analysis to determine whether they are at risk of contracting a certain TBD. The app was able to accurately identify three tick species: *Ixodes scapularis* (Eastern blacklegged tick), *Amblyomma americanum* (Lone star tick), and *Dermacentor variabilis* (American dog tick). The overall accuracy rate of the model is approximately 80%. This app shows promise in assisting tick bite victims with their decision of seeking further medical assistance, particularly those with underlying health conditions.

I. INTRODUCTION

More than one million people are bitten by ticks each year in the United States [1]. The Centers for Disease Control and Prevention report that 50,000 people are diagnosed with tick-borne diseases (TBD) each year [2]. Moreover, underreporting discovered by various studies [3] suggests that the real incidence rate is much higher, closer to ~500,000. Tick exposure and bites are also associated with other diseases, such as Pediatric Autoimmune Encephalitis Associated with Streptococcal Infections (PANDAS), Basal Ganglia Encephalitis (BGE), and other encephalopathies [4, 5]. In some cases, tick bites can lead to TBDs, such as Anaplasmosis, Babesiosis, Bartonellosis, Lyme Disease (LD), Chronic Lyme Disease (CLD), Post-Treatment Lyme Disease (PTLD), Ehrlichiosis, Tick-Borne Relapsing Fever (TBRF), Tularemia, Mycoplasmosis, Rocky Mountain Spotted Fever (RMSF), Powassan virus, and Southern Tick-Associated Rash Illness (STARI) [6, 7]. Despite existing antibiotic treatments, diseases such as LD and PTLD can become chronic if not treated immediately [8]. Moreover, LD, CLD, and PTLD account for more than 1.2 million affected people in the U.S. [9].

Traditional diagnostic methods have used species classification of ticks to inform whether a person is at risk of contracting a TBD, primarily LD. However, there is no direct correlation between a certain species of tick and the TBD that the species carries, since the risk factors vary greatly for both tick lifecycle and geographic location. Clinical studies of patients presenting with a tick bite in Lyme-endemic areas have shown that the prevalence of LD in all species of ticks range from 0% to 50% of the time

[10]. This includes the notorious biting arachnid Eastern blacklegged tick. With this degree of uncertainty, it is necessary to couple real-time identification with location-specific incidence rate data. Connecting these datasets in one platform is beneficial, since the general public and healthcare professionals alike find tick recognition challenging due to the large number of different species known [7]. There are 84 species of ticks in the U.S. alone, with significant variations in sizes and colors within each species (Figure 1). In fact, one study has shown that when a group of medical practitioners was asked to identify *I. scapularis*, one of the main carriers of LD, the accuracy rates were as low as 10.5% [11]. Correct identification requires in-depth entomological training. Given the limited knowledge in this field and the pressing health concerns worldwide over TBDs, deep learning-based models hold promise for the instantaneous assessment of TBD risks



based on specimen determination and prevalence analytics.

Figure 1. Sample selection of photos of female and male *A. americanum*, *I. scapularis*, and *D. variabilis* ticks that were used for training. Varying distinctive features of each species and sex are shown.

II. METHODS

Convolutional Neural Network (CNN) and Embedded Heuristics / Risks Assessment

Google Images, Global Biodiversity Information Facility, and iNaturalist were used to collect images of three species of ticks: *I. scapularis*, *A. americanum*, and *D. variabilis* (Figure 1). At least 200 images per species were used to train the CNN (600 total) and 150 additional images per species were used to test the CNN (450 total; 1,050 overall). Each image contained only one tick. Equal distributions of both sexes of ticks as well as a balance of engorged ticks (those that have ingested a large amount of blood from an animal or human and have a globular shape) were used during training and testing for each species. After comprehensive cleaning, the photos were used to train a CNN with a transfer learning approach and stochastic gradient descent as the optimization algorithm implemented in Python, open-source Keras, and TensorFlow 2.5.0 [12]. Keras InceptionV3 was utilized for the network's architecture, a method widely used for image classification [12]. Other network details for a total of 21,808,931 (21,774,499 trainable and 34,432 non-trainable) parameters were as follows: learning rate, 0.01; momentum, 0; weight initialization, ImageNet; class weights, equally balanced; and loss metric, categorical cross-entropy. After images were collected, they were closely examined to ensure that there was no grouping of other species within the searches or when the photos were downloaded. All photos were cropped to 160 × 160 px with Keras to enhance the training accuracy of the model. Out-of-focus images or photos with indistinguishable ticks were discarded from the dataset.

The tick classifier was then linked to the Pennsylvania Tick Research Lab’s Surveillance Program (TSP). This study has tested 64,970 ticks (as of 9/25/22) in the U.S. and other territories, and each state was given a categorical risk: deer tick or non-deer tick. Deer ticks are carriers of Lyme Disease, while non-deer ticks (ie: dog ticks) are often about double the size of deer ticks and are carriers of less known (but non-LD) pathogens. Of note, *I. scapularis* was categorized as a deer tick, whereas *A. americanum* and *D. variabilis* were classified as non-deer ticks. A function to provide the tick risk was built using Google Colab(atory) calling the TSP. Based on the geographic location from where a tick photo was uploaded, incidence rate surveillance data from the TSP were accessed live to determine the disease risk associated with a given tick found in that particular state. Notably, the TSP only reports the incidence rates of Babesiosis and Mycoplasmosis found in deer ticks (*I. scapularis*); RMSF, Tularemia, Ehrlichiosis, and STARI found in non-deer ticks (*A. americanum* and *D. variabilis*). LD, Anaplasmosis, Powassan virus, Bartonellosis, and TBRF found in both types of ticks in each state.

iOS Swift Development

DETICKT IT was developed in Xcode with Swift and within a service-oriented architecture. After the CNN model (Figure 2) was ported from a local environment to Amazon Web Services (AWS), open application programming interfaces were created for functions, such as location finder, risk assessment, and specimen identification. Using this approach, DETICKT IT is able to scale through AWS

Lambda, delivering results to the app’s users in mere seconds.

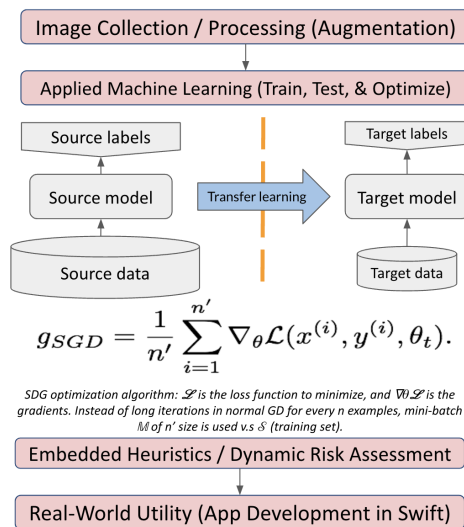


Figure 2. **Schematic approach for DETICKT IT.** DETICKT IT’s CNN uses transfer learning and embedded heuristics / risk assessment implemented in Python / TensorFlow, while the iOS app itself is coded in Xcode / Swift.

III. RESULTS AND DISCUSSION

CNN training and validation accuracy, which is the proportion of correct predictions used to validate the generalization ability of the model, increased gradually across training epochs. It also performed well with a

validation accuracy of 99.71% (Figure 3), using k-fold cross-validation with split training, testing, validation sets and a well-balanced distribution of the underlying data. While low bias and high variance may cause overfitting, high bias and low variance, in contrast, may prevent the model from obtaining sufficiently low errors in the training data (underfitting). Specifically, training the model for more epochs would result in low bias and high variance, while, by contrast, training it for fewer epochs would produce a high bias and low variance. The model’s complexity was effectively altered using k-fold cross-validation for regularization – with the overall accuracy of the model being over 85%.

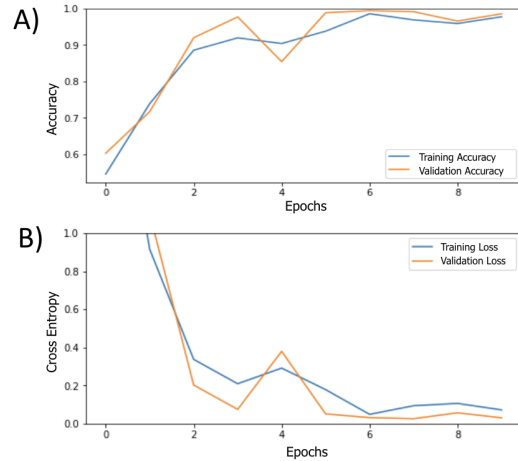


Figure 3. **Accuracy and loss metrics for the tick identification CNN.** Plots of the training (blue) and validation (orange) accuracies (A) and loss metrics (B) for the CNN are shown. n= 10 epochs.

Users can open DETICKT IT, take a photo of the tick they encounter, and instantly receive a comprehensive risk (potentiality) assessment of possible TBDs given the geo-locational heuristics paired with disease incidence rates derived from TSP (Figure 4). This is done based on the photo’s geotag, which identifies the specific location of where the photo was taken or where the tick was found.



Figure 4. **DETICKT IT App User Interface (UI).** (A) Home / Launch screen, (B) camera feature to take a photo of a tick, (C) tick identity based on species determined, and (D) full location-based risk assessment given the geotag from a photo of the *I. Scapularis* or the blacklegged tick.

In conclusion, this mobile app has the potential to significantly outperform the reported ability of medical professionals to accurately classify ticks [11]. Moreover, the embedded tick risk assessment feature allows users to gain

awareness of the risks associated with a given tick bite in a certain geographic area, in real time. The knowledge and confidence afforded by this easy-to-use app will facilitate medical professionals' diagnostic accuracy and shorten critical time-to-treatment for tick bite victims. This app has the potential to provide a better quality of life for affected individuals.

Going forward, DETICKT IT will be trained to identify additional species of ticks and extend to other countries for geography-specific risk assessment, and the app will also give users the ability to upload photos of ticks taken outside the app as well. Moreover, there will also be clinical resources for suggesting a compound risk score for persons bitten by a tick and identified by- or crowd-screened via DETICKT IT.

ACKNOWLEDGMENTS

We would like to acknowledge the King School Advanced Mathematics and Science Study Program Fund for funding. We would also like to thank Marcelo Busico for programming assistance.

REFERENCES

- [1] Stricker, R. B., & Johnson, L. (2014). Lyme disease: Call for a "manhattan project" to combat the epidemic. *PLoS Pathogens*, 10(1). <https://doi.org/10.1371/journal.ppat.1003796>
- [2] Marx, G. E., Spillane, M., Beck, A., Stein, Z., Powell, A. K., & Hinckley, A. F. (2021). Emergency department visits for Tick Bites — United States, January 2017–December 2019. *MMWR. Morbidity and Mortality Weekly Report*, 70(17), 612–616. <https://doi.org/10.15585/mmwr.mm7017a2>
- [3] Nelson, C. A., Saha, S., Kugeler, K. J., Delorey, M. J., Shankar M.B., Hinckley, A. F., & Mead, P. S. (2015). Incidence of clinician-diagnosed Lyme Disease, United States, 2005–2010. *Emerging Infectious Diseases*, 21(9), 1625–1631. <https://doi.org/10.3201/eid2109.150417>
- [4] Cross, A., Bouboulis, D., Shimasaki, C., & Jones, C. R. (2021). Case report: PANDAS and persistent Lyme Disease with neuropsychiatric symptoms: Treatment, resolution, and recovery, *front. Psychiatry*, 12(505941). [10.3389/fpsy.2021.505941](https://doi.org/10.3389/fpsy.2021.505941)
- [5] Bogovic, P., & Strle, F. (2015). "Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management," *World J. Clin Cases*, 3(5), 430–441. [10.12998/wjcc.v3.i5.430](https://doi.org/10.12998/wjcc.v3.i5.430)
- [6] Madison-Antenucci, S., Kramer, L. D., Gebhardt, L. L., & Kauffman, E. (2020). Emerging tick-borne diseases. *Clinical Microbiology Reviews*, 33(2). <https://doi.org/10.1128/cmr.00083-18>
- [7] Eisen, R. J., Kugeler, K. J., Eisen, L., Beard, C. B., & Paddock, C. D. (2017). Tick-borne zoonoses in the United States: Persistent and emerging threats to human health. *ILAR Journal*, 58(3), 319–335. <https://doi.org/10.1093/ilar/ilx005>
- [8] Stricker, R., & Johnson, L. (2016). Lyme disease: The promise of big data, companion diagnostics and precision medicine. *Infection and Drug Resistance*, Volume 9, 215–219. <https://doi.org/10.2147/idr.s114770>
- [9] Schwartz, A. M., Kugeler, K. J., Nelson, C. A., Marx, G. E., & Hinckley, A. F. (2021). Use of commercial claims data for evaluating trends in lyme disease diagnoses, United States, 2010–2018. *Emerging Infectious Diseases*, 27(2), 499–507. <https://doi.org/10.3201/eid2702.202728>
- [10] Shapiro, E. D. (2014). *Borrelia burgdorferi* (Lyme disease). *Pediatrics in Review*, 35(12):500-9. doi: 10.1542/pir.35-12-500. PMID: 25452659; PMCID: PMC5029759
- [11] Kopsco, H. L., Duhaime, R. J., & Mather, T. N. (2020). Assessing public tick identification ability and tick bite riskiness using passive photograph-based crowdsourced tick surveillance. *Journal of Medical Entomology*, 58(2), 837–846. <https://doi.org/10.1093/jme/tjaa196>
- [12] Hussain, M., Bird, J. J., & Faria, D. R. (2018). A study on CNN transfer learning for image classification. *Advances in Intelligent Systems and Computing*, 191–202. https://doi.org/10.1007/978-3-319-97982-3_16