CMPE257

Bug Bite Classification

Initial Report

Group 6

Clifford Lin - 017554069

Tin Pho - 014901003

Danishbir Singh Bhatti - 017521647

Sudip Das - 017549415

Suhaas Teja Vijjagiri - 016197935

Date: Feb 22, 2024

1. Idea/Solution Presentation

1.1. Context/Background

Insect bites are a common yet often overlooked health concern that can affect anyone at any time. The consequences of these bites extend beyond mere annoyance, as they can lead to skin reactions like redness, swelling, and other dermatological symptoms such as papules and urticaria [1]. Therefore understanding the types of bugs responsible for these bites and effectively classifying them is crucial for timely and accurate treatment.

People are aware of common bugs like spiders, mosquitos or ants, however they fail to realize how diverse each species is and what are the poisonous types in it. According to the national library of medicine [1], the phrase ‘bug bites’ is best defined as bites and stings inflicted by the member of the phylum Arthropoda. The fact that arthropods form the largest section of the animal kingdom (80% of all known animals), points out further on the size and diversity of bug bites. Hence narrowing it down to a specific region is vital for effective classification of the bite mark.

California has a diverse ecosystem harboring a multitude of insect species. Its varied landscape, ranging from coastal areas to mountainous regions and deserts, provides a habitat for a wide array of insects, including mosquitoes, ticks, spiders, and various biting flies. Due to the low visibility of the insects and the insignificant instant impact of the bite, it is generally ignored until it becomes a concern.

Arthropods are known to be vectors for a wide range of diseases caused by bacteria, viruses, and protozoa, posing a significant public health risk. For instance, kissing bugs can transmit the Trypanosoma cruzi parasite, which causes Chagas disease. This disease can lead to severe cardiac or gastrointestinal complications in 20 to 30% of those infected [2]. Another severe issue in the California area is mosquito-borne diseases such as West Nile virus. Individuals with weakened immune systems and older adults are vulnerable to the disease’s fatal effects including neurological illnesses.

The complexity of insect bites, combined with a general lack of knowledge about their sources, often leads to these incidents being ignored. In the United States, emergency department visits for arthropod bites and stings have shown recent trends, indicating the prevalence of this issue. Additionally, the prevalence of allergies in the U.S. population, including those caused by insect stings, further emphasizes the need for attention to this matter. It is estimated that between 0.4-0.8% of children and 3% of adults in the United States experience severe life-threatening systemic reactions from insect stings [3].

The development of a machine learning model for insect bite classification offers a promising approach to addressing this issue. By using convolutional neural networks (CNN) and deep learning algorithms insect bites can be identified with a higher accuracy. Therefore, through this project, we aim to reduce the incidence of misdiagnosis and provide immediate and reliable information to improve outcomes for those affected by these common yet potentially serious injuries.

1.2. Solution Initial Description

Firstly, in order to collect image data for our project, we will first investigate reputable data platforms to get the bug bite image datasets. However, this method is expected to contain biases in geographic diversity as the bug bite images may not appear only in California. Another method that we will do is contacting the entomology department in the universities in California to get more datasets. Secondly, the datasets will be carefully preprocessed by resizing and squishing it before using them to train the model. Thirdly, we will split the datasets to obtain training and testing datasets. Then, we will develop our main model using a convolutional neural network (CNN), which will detect the bites that appear on the users' skin. Finally, choosing features is essential as it will improve generalization, boost computational efficiency, and easily filter noisy data. In order to do this, we will choose the features to train the model through the filters, wrappers, and embedded methods, which are the commonly used techniques in image classification [4]. As we are focusing on small details that appear in the photos, we will evaluate the model using a quadratic relative error cost function (REMSE), which is evaluated to be more sensitive than the Mean Square Error cost function (MSE) to small anomaly elements due to its focus on relative error [5].

To classify bug bites from images, we have decided to adopt and utilize a convolutional neural network model. We do not use a logistic regression model as convolutional neural networks are more suitable for this task [6]. By using CNN and utilizing hidden layers such as Convolutional Kernels, MaxPooling, and ReLu, we are able to reduce the number of input nodes, observe similarities in complex images, and also learn positional differences between similar images [7]. By experimenting with different hidden layers, different activation functions, and different neural network depths, CNN provides more functionality than the logistic regression model while allowing a higher ceiling in accuracy for this image classification task.

Furthermore, to classify bug bites using additional metadata such as location and time after bite, we can simply concatenate the metadata with our last flattened convolutional layer of image classification. The convolutional neural networks allow for this type of flexibility in classification tasks. It adds a layer of complexity to our model but hopefully will yield a higher accuracy in determining the bug bite [8].

1.3. Goals/Objectives

Our goal is to classify and detect the bug bites and the effects on the skin. We have found 2 pre-existing models that are following a similar approach to bug-bite identification[10][11]. Our objective is to increase the accuracy of our model and make it better, efficient and more reliable than the pre-existing ones. Our goal also includes some other metrics to increase its functionality that will include the meta data acquired from the people and dataset which can help us to identify how the different places and environmental conditions change the outcome or effect of a bug-bite. We also want our project to help the people in the long-run by making it a real-life problem solution and if possible, collaborating with the health sector to enhance and increase functionality to a next level. Our future aspects of the project aims for the betterment of people by introducing it on a much smaller level (for eg. Bay Area) and by increasing its accuracy to near perfect, after that, we also want to expand its boundaries at a much larger scale.

2. Project Management

2.1. Group Presentation

2.1.1. Member Short Bio/Introduction

- Tin Pho: I am an international student from Vietnam. I finished my Bachelor's degree in Software Engineering from San Jose State University in August 2023. I joined the Artificial Intelligence Master program because I want to become an AI engineer.
- Clifford Lin: Hi, my name is Cifford and I'm currently a masters student at San Jose State University studying artificial intelligence. I want to use this project to gain the relevant knowledge to become a ML engineer.
- Danishbir Singh Bhatti: Hey, I am Danishbir Singh Bhatti and I finished my bachelors in ECE. I have always been interested in ML and AI, and also did many projects in Machine Learning throughout my undergraduate degree just by self-learning.
- Sudip Das: Hey, I am Sudip and I am currently pursuing my Master's degree in Artificial Intelligence. My main motivation is to explore the fields of AI and ML, and ultimately be a part of the new innovations.
- Suhaas Teja Vijjagiri: Hey, I am Suhaas and I am currently pursuing my Master's degree in Computer Engineering. I like tech and I want to learn more about Machine Learning.

2.2. Planning

2.2.1. Task Distribution (Gantt Chart)

Gantt Chart (1)

2.2.2. Project Milestones

- Milestone 1: Dataset Acquisition
  - Description: Collecting qualitative data through online platforms and entomology departments from universities. The datasets have to have multiple skin tones and a variety of different bug bites that appear in California.
  - Estimated Time Frame: 02/23 - 03/08
- Milestone 2: Data Preprocessing
  - Description: Cleaning, normalizing, and splitting the datasets thoroughly.
  - Estimated Time Frame: 03/09 - 03/16
- Milestone 3: Model Development
  - Description: Designing CNN bite detection model. The goal is to recognize the skin areas that are bitten and give a prediction of the bugs using possibility.
  - Estimated Time Frame: 03/17 - 04/08
- Milestone 4: Training the Model with Feature Selection, Optimization and Model Evaluation

- Description: Using filters, wrappers, and embedded methods to get the most impactful features and avoid underfitting or overfitting and evaluating model performance using cost function.
- Estimated Time Frame: 04/09 - 04/16

2.2.3. Group Communication Tools & Techniques

Our team will collaborate with each other in-person as well as digital platforms like Zoom and Discord. We are also planning to do in-person brainstorming sessions which focus on efficient problem solving techniques. This will allow us to have real-time discussions, feedback and file-sharing ensuring that our team remains well-connected and well-coordinated despite the physical constraints.

2.2.4. Project Tools/Frameworks

One of the most important tools we need for this project is a large dataset of labeled images of different bug bites. Though we can collect images online, it would not provide enough metadata such as location or time after bite for these images. Thus we plan on contacting various Entomology departments to utilize their image dataset on bug bites. Other sources of our dataset include Kaggle and MDPI. We will be using Github to collaborate asynchronously while using Python and Jupyter Notebook to prepare the model. Some frameworks we plan on using include Numpy for math operations, Pandas for data manipulation, OpenCV and Pillow for image processing, and Keras for neural network implementation and classification [7]. If we take this project one step further, we will integrate our model with a web application for general use when it is complete. It will be created using HTML, CSS, Javascript and other web development frameworks such as React.js.

References

[1] J. Powers, "Insect Bites," Continuing Education Activity, [Online]. Available: https://www.statpearls.com/ArticleLibrary/viewarticle/129407. [Accessed: 18-Feb-2024].

[2] "Kissing Bugs," StatPearls, NCBI Bookshelf, [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK538239/. [Accessed: 18-Feb-2024].

[3] "Arthropod Bites and Stings," AAFP, [Online]. Available: https://www.aafp.org/afp/2003/0701/p119.html. [Accessed: 18-Feb-2024].

[4] N. Zhang, S.-L. Shen, A. Zhou, and Y.-S. Xu, "Investigation on Performance of Neural Networks Using Quadratic Relative Error Cost Function," in IEEE Access, vol. 7, pp. 106642-106652, 2019, doi: 10.1109/ACCESS.2019.2930520.

[5] Bolon-Canedo, Verónica, and Beatriz Remeseiro. "Feature selection in image analysis: a survey." Universidad de Oviedo, 2013, [Online]. Available: https://digibuo.uniovi.es/dspace/bitstream/handle/10651/53052/vbolon_aire2020.pdf?sequence=1 . [Accessed: 18-Feb-2024].

[6] S. Barari, "Deep Learning in Python: Creating a Convolutional Neural Network," Jan. 2019, doi: https://doi.org/10.4135/9781526493460.

[7] K. Kalra, "Convolutional Neural Networks for Image Classification," *Medium*, Jul. 14, 2023. https://medium.com/@khwabkalra1/convolutional-neural-networks-for-image-classification-f0754f7b94aa

[8] A. V. Akshaykrishnan, C. Sharanya, K. Abhinav, K. C. Aparna, and P. V. Bindu, "A Machine Learning based Insect Bite Classification," 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), 2023. [Accessed: 18-Feb-2024].

[9] J. S. Ellen, C. A. Graff, and M. D. Ohman, "Improving plankton image classification using context metadata," *Limnology and Oceanography: Methods*, vol. 17, no. 8, pp. 439–461, Jul. 2019, doi: https://doi.org/10.1002/lom3.10324.

[10] S. Han, C. Buonocore, "Bug Bite AI", 2019, Devpost, [Online]. Available: https://devpost.com/software/bugbite-tf

[11] Ilijoski, B.; Trojachanec Dineva, K.; Tojtovska Ribarski, B.; Petrov, P.; Mladenovska, T.; Trajanoska, M.; Gjorshoska, I.; Lameski, P. Deep Learning Methods for Bug Bite Classification: An End-to-End System. *Appl. Sci.* **2023**, *13*, 5187. https://doi.org/10.3390/app13085187