**Introduction**

Snakes are among the top three most dangerous creatures, and identification is very difficult, often relying on strong herpetological skills that require identification by characteristics such as head shape and body pattern or colour[3]. This approach has its limitations, and by automating the identification process, it will be easier for people to avoid venomous snakes while also helping healthcare providers provide better treatment.

The use of ML algorithms has the potential to greatly increase the accuracy of snake identification, and these algorithms have the potential to mitigate the negative impacts of mistreatment of snake bites. It is very important to understand whether a given snake is venomous or not since diagnosis and treatment differ greatly between the two[1]. Studies have found that 12% of nonvenomous snakebites were treated as if there were necrosis[2], and this number can be greatly decreased by these algorithms.

There are other benefits in the healthcare industry as well. These algorithms have the potential to be a low-cost alternative to having skilled individuals classify snakes, helping to provide aid to healthcare providers in low-resource settings. Snake venom is also a sought-after drug, and classifying a snake could help scientists quickly find which snakes are useful for harvesting venom[2]. Furthermore, efficient classification could also help zoologists and conservationists alike better understand snake populations around the globe.

**Problem Definition**

Accurate classification of snakes is important for identifying how venomous or nonvenomous the creature is to humans. The goal of this project is to identify snake species quickly and accurately, minimising false negatives and false positives to ensure accurate identification.

Our dataset has 135 different species of snakes using 24,000 images from a Kaggle dataset. Each image is labelled by the binomial name for the snake, the country where it is found, the continent, genus, family, and sub-family

**Methods**

We are planning on using CNN, Decision Tree, and K-Means Algorithm (scikit-learn) to accomplish this task. CNN is a traditional algorithm used for image classification, and using it will likely yield the most accurate results. It will be interesting to compare the results with the K-means algorithm and see how it compares with CNN, as it’s a much simpler clustering algorithm that can be used to divide the dataset into different categories based on species.

**Potential Results, Metrics**

A confusion matrix will be used to evaluate the classification accuracy of a given algorithm, and precision-recall curves will be used to measure the false positive and false negative rates of these different algorithms. Along with this, we will use the F1 score to determine how accurate each model is.

**Checkpoint**

We now have a problem and the motivation for it. We have our dataset and we have three methods we will use to analyse the dataset and train our model – These methods and metrics allow us to compare how the different algorithms compare when solving the same problem, and we are ready to move on to preparing the data and using it to train our model.

References

[1] Niteesh., I., Venkat.A, M. S., Vahed., S., Dattu.P, N., & Srilatha., M. (2021). Classification and prediction of snake species based on snakes’ visual features using machine learning. 2021 2nd Global Conference for Advancement in Technology (GCAT). https://doi.org/10.1109/gcat52182.2021.9587711

[2] Progga, N. I., Rezoana, N., Hossain, M. S., Islam, R. U., & Andersson, K. (2021). A CNN based model for venomous and non-venomous snake classification. Applied Intelligence and Informatics, 216–231. https://doi.org/10.1007/978-3-030-82269-9\_17

[3] Rajabizadeh, M., & Rezghi, M. (2021). A comparative study on image-based snake identification using machine learning. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-96031-1

Contribution Table:

|  |  |
| --- | --- |
| Jaden Co | * Discussing potential results and performance metrics * Creating GitHub Repository * Recording audio for proposal video presentation |
| Robert Jeon | * Recording audio for proposal video presentation |
| Karan Patel | * Finding viable dataset * Recording audio for proposal video presentation |
| David Qu | * Finding references on the topic * Writing introduction and background, problem definition * Recording audio for proposal video presentation |
| Jehyeok Woo | * Recording audio for proposal video presentation * Helping to populate GitHub page |