**Snake Species identification using Machine Learning**

Mohd Aamir, Emerson Cardoso, Sameer Hans

EURECOM

MALIS project proposal

**Abstract:** *Snakebite poisoning is one of the tropical diseases responsible for many deaths globally and particularly in developing countries. In most cases, identifying the species of the snake following the bite is the life-saving factor for snakebite poisoning patients. However, snake species identification is not a trivial task, it can be misleading and dangerous. That is why this study proposes the use of Machine Learning approaches to automate snake species identification.*

# Introduction

Snakes are vertebrate animals that belong to the group of reptiles, among which approximately 700 species are described as poisonous. Snakes are animals that share physical characteristics between different species, have different characteristics than snakes of the same species in different geographic regions, the characteristics of species vary at different ages and in opposite sexes.

Traditionally the snake identification process is carried out by making a visual analysis. This approach proves to be inefficient, as it can be misleading due to the need for in-depth knowledge about these animals. For example, to identify correctly, one needs to be able to distinguish different patterns that can be found in a snake which may vary in age, location, sex, the climate in the region, among other features. Also, visual analysis can be dangerous as it needs certain physical contact or approximation between the snake and the specialist

Snakebite poisoning is a potentially fatal disease, causes about 81000 - 138000 deaths annually, and about 421000 and 1.2 million poisonings per year (Bolon et al., 2020). Moreover, the world health organization (WHO) has declared snakebite to be a neglected tropical disease (NTD) that causes a humanitarian crisis (de Castañeda et al., 2019; *Snakebite*, n.d.).

# Motivation

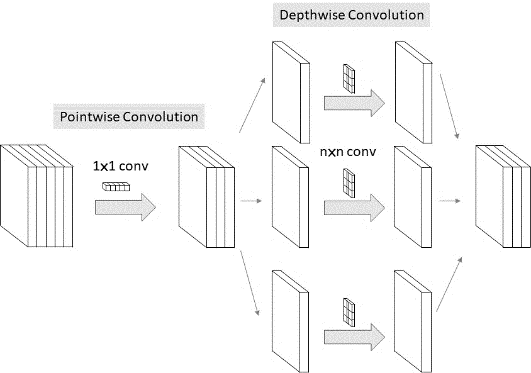
According to WHO the snakebit causes over 100,000 human deaths and 400,000 victims of disability and disfigurement globally every year. It affects poor and rural communities in developing countries, which host the highest venomous snake diversity and the highest-burden of snakebite due to limited medical expertise and access to antivenoms. Death or harmful effects of snakebites can be reversed or prevented by administering antidotes. However, administering an antidote is only efficient when done quickly and correctly -with a positive snake species classification.

# Methods

In this study, we will make a comparison between Convolutional Neural Network Xception and one method of machine learning(K-NN).

## Models

**Xception:** (Chollet, 2017) which means “eXtreme Inception”, is a linear architecture network of 36 layers organized in 14 modules that have a residual connection in each module except the first and last module. This architecture uses depth wise separable convolutions instead of Convolution Layer. A depthwise separable convolution consists in a depthwise convolution, which is a spatial convolution performed independently over each channel of an input, followed by a pointwise convolution, i.e. a 1x1 convolution, projecting the channel's output by the depthwise convolution onto a new channel space. The Xception used a modified depthwise separable convolution where the first layer perform a pointwise convolution followed by depthwise convolution as shown in figure 1.



**K-NN:** The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. K-NN - algorithm assumes that similar things exist in proximity. In other words, similar things are near to each other.

## Dataset

In this project we will be using part of the dataset released in the challenge proposed by the world health organization (WHO) for snack identification, more concrete in this study we will work with 8 species of venomous snake (table 1) from France and Europe in general. From the WHO dataset we were able to collect 1222 Images in total, where the species 000(Vipera aspis) have more image 247 in total and the class 005 (Vipera Xanthina) is the species with less image 22. To split our dataset in train and test we use the ration 80:20, where 20% of the images per class was attribute to test set. Before splitting the dataset in training and test we did some data normalization such as resize all the image to the same size 225x225.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Snake Species** | **Images** | **Train** | **Test** |
| *000* | Vipera aspis | 237 | 190 | 47 |
| *001* | Vipera berus | 280 | 224 | 56 |
| *002* | Vipera ursinii | 36 | 29 | 7 |
| *003* | Vipera seoanei | 93 | 74 | 19 |
| *004* | Vipera ammodytes | 119 | 95 | 24 |
| *005* | Vipera xanthina | 22 | 18 | 4 |
| *006* | Coronella girondica | 196 | 157 | 39 |
| *007* | Malpolon monspessulanus | 239 | 191 | 48 |
| **Total** | | **1222** | **978** | **244** |

The ratio between the class with lowest sample and the one with higher sample is xx:xx we can se clearly how despoportion is the dataset, for that reaseon we decide to use data augmentation on train set to provide well balanced dataset set. We aply data augmentation for class with few sample so that we can have the same number of sample with the class with more sample. The following techniques of data augmentation was used randonaly.

The final dataset has some interesting characteristics that make the classification of these species a challenge, some of these characteristics are described below:

The dataset contains a images of snakes in their habitat (bushes, rocks, etc.), which makes classifying these images difficult because some snakes…..

And according with … the snake are very diversity animal which makes somes species pattern (color, etc) be very different in species age, location and sex….

**Related works**

The table below shows some of the previous work related to snake identification

|  |  |  |
| --- | --- | --- |
| **Author** | **Approach** | **Accuracy (%)** |
| Aimir [4] | KNN (k=7) | **80.34** |
| Naïve Bayes | **75.64** |
| Decision tree J48 | **71.29** |
| Backpropagation neural networks | **87.93** |
| Nearest neighbour | **89.22** |
| (Abeysinghe et al., 2019) | Siamese | **5** |
| (Abdurrazaq et al., 2019) | Shallow CNN | **75.19** |
| Medium CNN | **82.14** |
| Deep CNN | **78.74** |

**4. Experiments and results**

In this study, we will train our model to identify 8 species of snake which can be found in the European of the world. To evaluate the performance of our models we will use K-fold cross-validation with K=10 and the following metrics will be computed in each fold, precision, accuracy, f1\_score and recall, we will compute the model performance as the average of these metrics.

Xception

1. **References**

Abdurrazaq, I. S., Suyanto, S., & Utama, D. Q. (2019). Image-Based Classification of Snake Species Using Convolutional Neural Network. *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, 97–102. https://doi.org/10.1109/ISRITI48646.2019.9034633

Abeysinghe, C., Welivita, A., & Perera, I. (2019). Snake image classification using siamese networks. *ACM International Conference Proceeding Series*, *June*, 8–12. https://doi.org/10.1145/3338472.3338476

Amir, A., Zahri, N. A. H., Yaakob, N., & Ahmad, R. B. (2017). *Image Classification for Snake Species Using Machine Learning Techniques*. https://doi.org/10.1007/978-3-319-48517-1

Bolon, I., Durso, A. M., Mesa, S. B., Ray, N., Alcoba, G., Chappuis, F., & De Castañeda, R. R. (2020). Identifying the snake: First scoping review on practices of communities and healthcare providers confronted with snakebite across the world. *PLoS ONE*, *15*(3), 1–24. https://doi.org/10.1371/journal.pone.0229989

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

de Castañeda, R. R., Durso, A. M., Ray, N., Fernández, J. L., Williams, D. J., Alcoba, G., Chappuis, F., Salathé, M., & Bolon, I. (2019). Snakebite and snake identification: empowering neglected communities and health-care providers with AI. *The Lancet Digital Health*, *1*(5), e202–e203. https://doi.org/10.1016/S2589-7500(19)30086-X

*Snakebite*. (n.d.). Retrieved February 26, 2020, from https://www.who.int/health-topics/snakebite#tab=tab\_3