

# Ancient Languages Classification using Artificial Neural Networks

<https://doi.org/10.63962/NOPF5551>

Safia Al Ali  
MSc in Business Analytics – AI Management  
Abu Dhabi School of Management  
Abu Dhabi, UAE  
safia4489@gmail.com

**Abstract**—The study of ancient languages is essential to gain a comprehensive understanding of mankind's history. Recently, there has been a growing interest in leveraging AI in this area. The aim of this paper is to apply AI deep learning classification models to five ancient languages which are Ancient Greek, Arabic, Old Chinese, Egyptian Hieroglyph, and Sanskrit. Three CNN Residual Network models were implemented: ResNet34, ResNet50, and ResNet101. Results revealed that ResNet101 is the most accurate model with 92% rate achieved. In addition, Sanskrit is the most accurately predicted language and Arabic is the least accurate. Further research suggests expanding the dataset, including more languages, and exploring other models.

**Keywords**—Ancient Languages, AI, Deep Learning, Convolutional Neural Networks, Residual Network

## I. INTRODUCTION

Ancient languages vary tremendously from their modern counterparts, presenting unique challenges and requiring significant effort and expert involvement. The aim of this thesis is to bridge the linguistic gap by harnessing the power of Neural Network Artificial Intelligence models to develop an advanced tool capable of classifying various ancient languages. A distinctive dataset is compiled specifically for this study. There are a few expected limitations to be considered, such as the requirement for a massive amount of data for optimal performance, which is not possible due to time constraints and CPU limitations. In addition, ancient languages have been written in various media and may have evolved over history which adds complexity and variation. Furthermore, model validation and accuracy might require archeologists' involvement in real world applications, therefore the evaluation of performance will be based on test data [4].

## II. METHODOLOGY

### A. Business Understanding

The chosen methodology, CRISP-DM, will be followed step by step to achieve the thesis objective of classifying various ancient languages. Below is a summary of main steps followed [6].

### B. Data Understanding

The collected dataset consists of images from various online public sources. This dataset is exclusively gathered with the aim of having different groups of images for at least five ancient languages divided into model testing, training, and evaluation. The dataset is stored locally. Unlike structured tabular format, image dataset is simple with limited variables such as language name, title of the image, its type, and size.

### C. Data Availability and Ethics Statement

The dataset used in this study was self-compiled by the author through the collection of publicly available images from Google Images, ensuring that only images licensed for reuse were selected. These images represent five different ancient languages and were curated specifically for training and evaluating the neural network model described in this paper. Since the dataset is composed of non-personal, publicly available images, and internal ethical approval was acquired for this research from our Research committee. The dataset is not hosted in a public repository, but it can be made available upon reasonable request to the corresponding author for academic and research purposes only.

### D. Data Preparation

To ensure optimal image dataset is collected and preprocessed, the following steps were accomplished:

- **Data cleaning:** While gathering data, it was noticed that certain images have modern languages titles as part of the image. This must be cropped to avoid disturbing the accuracy of running the machine learning algorithm later.
- **Data formatting:** The standard types of images are JPEG and PNG. However, there were unusual types such as WEBP which the model might not be able to read. Therefore, such images will be converted to JPEG or substituted with alternatives.
- **Data Selection:** The quality of certain images gathered were not clear. Such images also require replacement to avoid disrupting the accuracy of the model.

### E. Modeling

Modelling involves organizing the collected dataset to adhere to a predefined configuration setting. The configured parameters are recommended to establish a robust model and enhance performance. The collected dataset will be structured into three different main pre-labeled categories or folders which are: train, test, and valid. Each designated folder will include a curated collection of images dataset that will represent 5 distinct ancient languages which are: Egyptian Hieroglyphs, Chinese, Ancient Greek, Arabic, and Sanskrit. Thirty images per language will be allocated for the train folder to ensure sufficient and effective model training. The test folder will contain 10 images for each language, and another 10 images will be allocated to the valid folder. It's essential to allocate training, testing, and validation to ensure reliability and generalization [5]. A considerable portion of collected data will be allocated to the training; thus model can learn features and patterns. The chosen model for this case study is the Convolutional Neural Network (CNN). CNN models are designed specifically for images data for tasks such as classification. Docker containers and Kubernetes will be launched to run the Python code of the chosen CNN models: ResNet34, ResNet50, and ResNet101.

### F. Evaluation

When implementing CNN models, several evaluation methods can be applied to assess the algorithms. In this thesis, the main matrices that will be used are the confusion matrix, prediction/actual values, and loss/probability scores.

### G. Deployment

Successful implementation depends not only on model accuracy but also on its usability and accessibility. There are various methods of deployment strategies and in this thesis paper we propose the following: tool integration in different platforms such as websites, mobile applications, or systems that are specialized in linguistic and historical fields. Examples of such tools can be ancient language translation and archives [3]. Another deployment recommendation is to include the model in the GitHub platform. GitHub is a well-known advanced platform that allows collaboration, sharing, and machine learning deployment [2]. Stand-alone downloadable python files are another useful approach of deployment.

## III. DATA ANALYSIS

There are various insightful techniques to conduct image dataset analysis, this section includes some of these techniques which were applied for the ancient language classification model.

### A. Visual Inspection

Visual Inspection of an image can provide various details such as image quality, condition of the manuscript, paper and ink, medium and craftsmanship, condition of the surface, resolution, clarity, and visual consistency.

### B. Statistical Analysis

To perform statistical analysis of image dataset, specific codes were run in Python Google Colab. Examples of statistical analysis are image size or dimension, color distribution, edge detection, pixel intensity distribution, and texture analysis.

## IV. RESULTS

Three CNN models will be used for images dataset classification task of ancient languages. The breakdown of the Python code steps and the achieved results is as below:

- The code will start by importing the required libraries and setting up the data using the Image Data Bunch function from the specified data path uploaded.
- Data exploration phase will also be included at this stage to explore and identify different variables such as classes names and sample batch display.
- 3 CNN models will be run which are: ResNet34 Model, ResNet50 Model, and ResNet101 Model

### A. ResNet34 Model

The first chosen CNN model is ResNet34 which is implemented in Python by creating the CNN function and training of 4 epochs. Epochs is the frequency which the algorithm goes over the whole training dataset to learn [1]. The initial result of the model achieved **52%** performance accuracy.

The confusion matrix and most confused classes (languages) are displayed by adding the classification interpretation function. in Python. Confusion Matrix and most confused graphs revealed that model confused Arabic language with Sanskrit 7 times, Sanskrit with Old Chinese 5 times, and both Ancient Greek language and Egyptian with Arabic 3 times. While the model was able to accurately classify Sanskrit language at all 10 times of validation process.

Then the ResNet34 model was unfrozen to go through fine tuning process with further training for one more epoch which resulted in slightly higher accuracy of **54%** as illustrated in figure 1.



Fig. 1. ResNet34 Fine Tune Process Result

### B. ResNet50 Model

CNN ResNet50 model was the second model applied with an increase image size to 299 and smaller batch size. The size of batch indicates the number of sample data handled by the model before the update occurs [1].

After training of 8 epochs, the achieved accuracy of validation image dataset reached **84%**. Fine tuning process was applied to unfreeze the model and train for 6 times more epochs which enhanced the performance accuracy to achieve **88%** as shown in Figure 2.

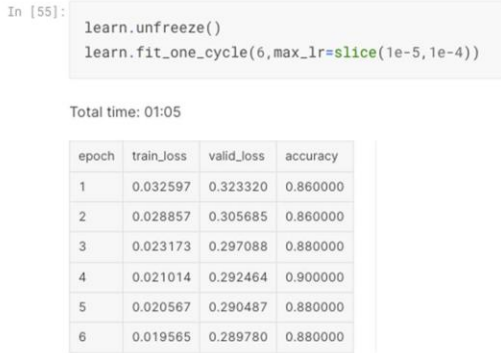


Fig. 2. Fine Tune Process ResNet50

Most confused graphs revealed that model confused Arabic language with Sanskrit 2 times and confused each of the following one time only: Ancient Greek language with Sanskrit, Arabic with Egyptian, Old Chinese with Egyptian, and Egyptian with Greek.

Confusion Matrix revealed that Sanskrit is the most accurate language with all 10-validation dataset predicted accurately, followed by Old Chinese, Egyptian Hieroglyphs, and Ancient Greek with each 9-validation dataset accurately predicted. The least accurate language is Arabic with only 7 validated data points correctly predicted.

### C. ResNet101 Model

The last CNN model applied is ResNet101 model with an increase image size to 384. Similar to ResNet50, 8 epochs were trained, the achieved performance accuracy of validation dataset is **88%**.

The model went through fine tuning process to train for 6 more epochs which improved performance accuracy to achieve **92%** as revealed in figure 3.



Fig. 3. Fine Tune Result of ResNet101

Most confused graphs revealed that model confused one-time Arabic language with each Sanskrit, Egyptian, Ancient

Greek. In addition, it confused Ancient Greek with Old Chinese.

Confusion Matrix revealed that Sanskrit, Old Chinese, and Egyptian are accurately predicted languages with all 10-validation dataset, followed by Ancient Greek with 9-validation dataset accurately predicted. The least accurate language is still Arabic with 7 correctly predicted as illustrated in figure 4.

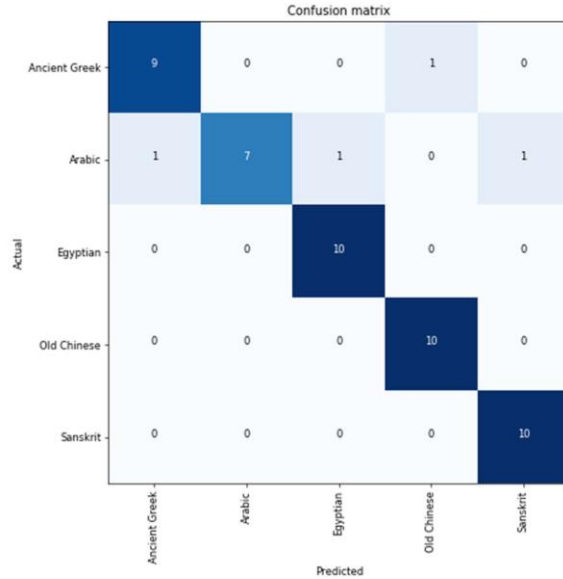


Fig. 4. ResNet101 Confusion Matrix

To summarize the results, table 1 and table 2 illustrate and compare how implemented ResNet CNN models architecture depth and image size play a significant role in the performance of classification task for ancient languages. In general, as the model depth and image size increased, the performance accuracy improved and confusion between languages decreased noticeably. Fine-tuning played a vital role in enhancing the accuracy of performance by leveraging pre-trained features specially when dealing with special domain such as ancient languages classification. ResNet101, which is the model with profound depth and largest image size, is recommended as it achieved highest performance accuracy of 92% after fine tune process were implemented as shown in table 1. In addition, RestNet101 achieved the lowest error rate of just 6% with 46 out 50 values predicted accurately as revealed in table 2.

CNN Model	Image Size	Accuracy	Fine Tune Accuracy
ResNet34	224	52%	54%
ResNet50	299	84%	88%
ResNet101	384	88%	92%

TABLE I. RESNET MODELS PERFORMANCE RESULTS

CNN Model	Error rate	Number of correct predicted values	Most Confused Language	Most Accurate Language
ResNet34	40%	30	Arabic	Sanskrit
ResNet50	12%	44	Arabic	Sanskrit
ResNet101	6%	46	Arabic	Sanskrit, Old Chinese, Egyptian

TABLE II. RESNET MODELS CONFUSION MATRIX RESULTS

## V. RECOMMENDATIONS AND CONCLUSION

Based on the revision, experiment, and output of the implemented classification model for the ancient languages task, it is recommended for future research to consider the following:

- Continuous experiments with different parameters for the three models and different image sizes.
- Focused scientific studies in most confused languages such as Arabic, which tend to be more complicated than others and require separate attention to its patterns and details.
- The total image dataset used for this thesis is 250 for 5 distinct languages. The number of collected dataset can be increased as well as number of languages to produce more profound solid ancient language classification models.
- The involvement of linguistic experts can be helpful as they will add useful insights.
- In addition, Other Convolutional Neural Network can be explored to understand if there are better models for ancient languages classification task.

In conclusion, understanding humanity's history, languages, cognitive development, and culture is the underlying

motivation behind researching ancient languages. However, there are many challenges in this field due to scarcity of linguistic materials and the extensive amount of time and expertise required. In recent decades, there has been an increased interest and scholarly attention in implementing advanced technological tools such as Artificial Intelligence and machine learning to overcome these challenges. Implementing AI and deep learning models for Ancient Languages classification task was the main driven to research and add value for linguistic experts with providing useful tool for their studies and work.

## REFERENCES

- [1] Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network. *Machine Learning Mastery*, 20.
- [2] Cosentino, V., Izquierdo, J. L. C., & Cabot, J. (2017). A systematic mapping study of software development with GitHub. *Ieee access*, 5, 7173-7192.
- [3] Dong, X. L., & Rekatsinas, T. (2018, May). Data integration and machine learning: A natural synergy. In *Proceedings of the 2018 international conference on management of data* (pp. 1645-1650).
- [4] Guidi, T., Python, L., Forasassi, M., Cucci, C., Franci, M., Argenti, F., & Barucci, A. (2023). Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks. *Algorithms*, 16(2), 79. <https://doi.org/10.3390/a16020079>
- [5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [6] Nama, G. F., Nabella, R. O., Wintoro, P. B., & Mulyani, Y. (2022). Towards CRISP-ML (Q): Consumer Segmentation Knowledge Analysis Using Data Science Approach for Marketing Strategy Recommendations. *NeuroQuantology*, 20(5), 3479-3489. [10.14704/nq.2022.20.5.NQ226471](https://doi.org/10.14704/nq.2022.20.5.NQ226471)