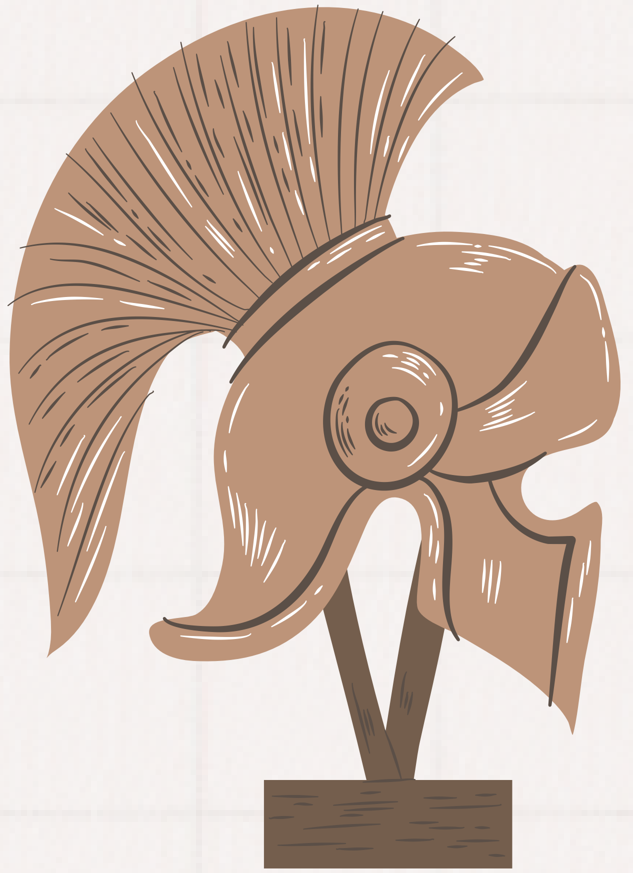
Welcome to our capstone project

# SIMILARITY DETECTION OF INDUS VALLEY SCRIPT WITH THE SCRIPTS OF VARIOUS OTHER INDIC LANGUAGES




**with our team**

Nikhil Kumar Saini- AP21110010002
Akash Ghosh- AP21110010057
Vrijeshwar Singh- AP2111010922
Durga Prasad Oleti- AP21110011

# TODAY WE'LL DISCUSS

- Motivation
- Objective
- Recent claims
- Dataset prep.
- Data overview
- Character recon.
- Feature ext.
- Training & eval.
- Results
- Conclusion& future

# MOTIVATION

- -The Indus Valley Civilization had a unique script, but no one has been able to fully read it yet.
- Many researchers, like Yagnadevam (Iravatham Mahadevan), believe it may be related to early Indian languages.
- Ancient scripts like Brahmi, Grantha, Kharosthi, and Brahui could hold clues to understanding the Indus script.
- Most current studies are based on manual analysis and historical records.
- With the help of Machine Learning, especially image-based models like CNNs, we can:
  - Analyze the shapes (glyphs) of characters in different scripts.
  - Find visual patterns and similarities that humans may miss.
- This could help support or challenge existing language theories about the Indus script.
- Our goal is not to decode the language completely, but to provide evidence of script relationships using visual similarity.

# OBJECTIVE

- To identify visual similarities between the Indus Valley Script and other ancient Indic scripts using Machine Learning.

- To use glyph (character) shapes as the basis for comparison since language may differ, but visual forms can offer clues.

- Extract visual features from characters using Convolutional Neural Networks (CNNs).

- Convert these features into numeric form (feature vectors) so they can be analyzed and compared.

- Train a model on known scripts and test how closely Indus glyphs match any of them.

- Support or question historical theories of Yagnadevam using a data-driven, statistical approach.

# RECENT CLAIMS ON IVC SCRIPT

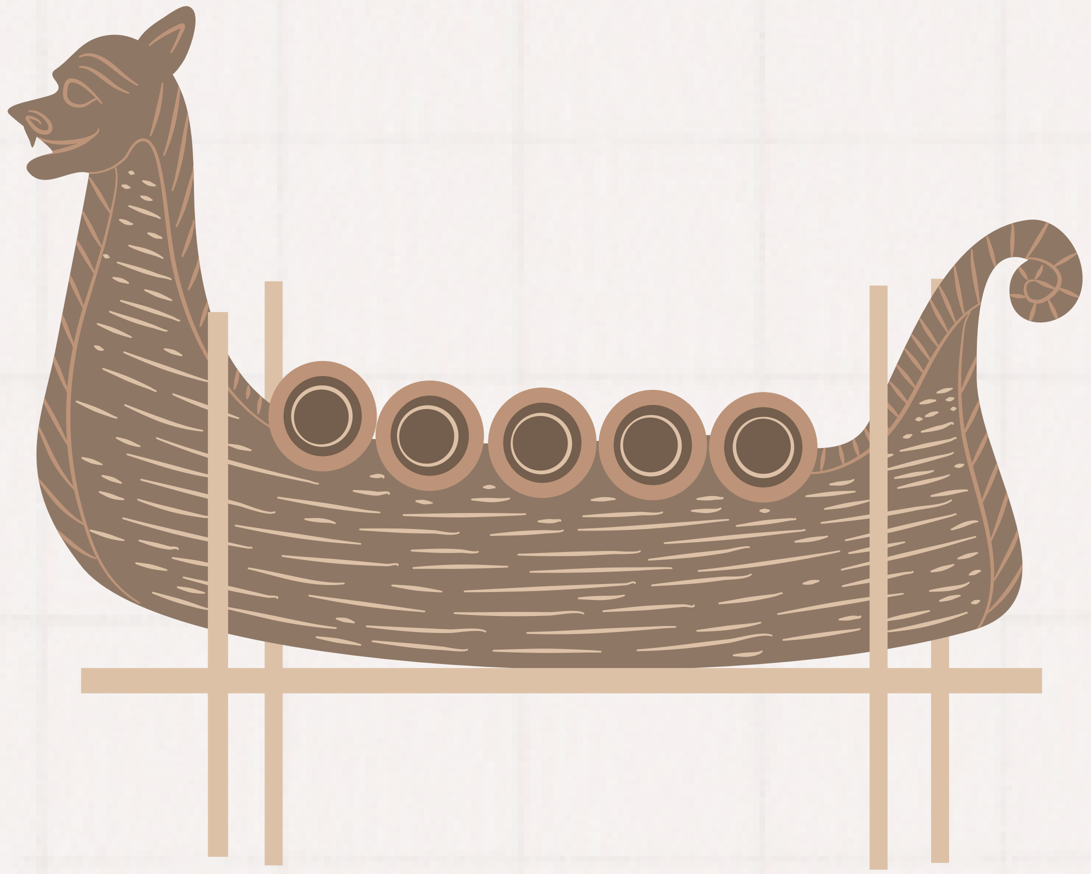
- Yajnadevam was a leading expert on the Indus script and strongly supported the Dravidian hypothesis, linking the script to early South Indian languages like Brahmi, Tamil-Brahmi.

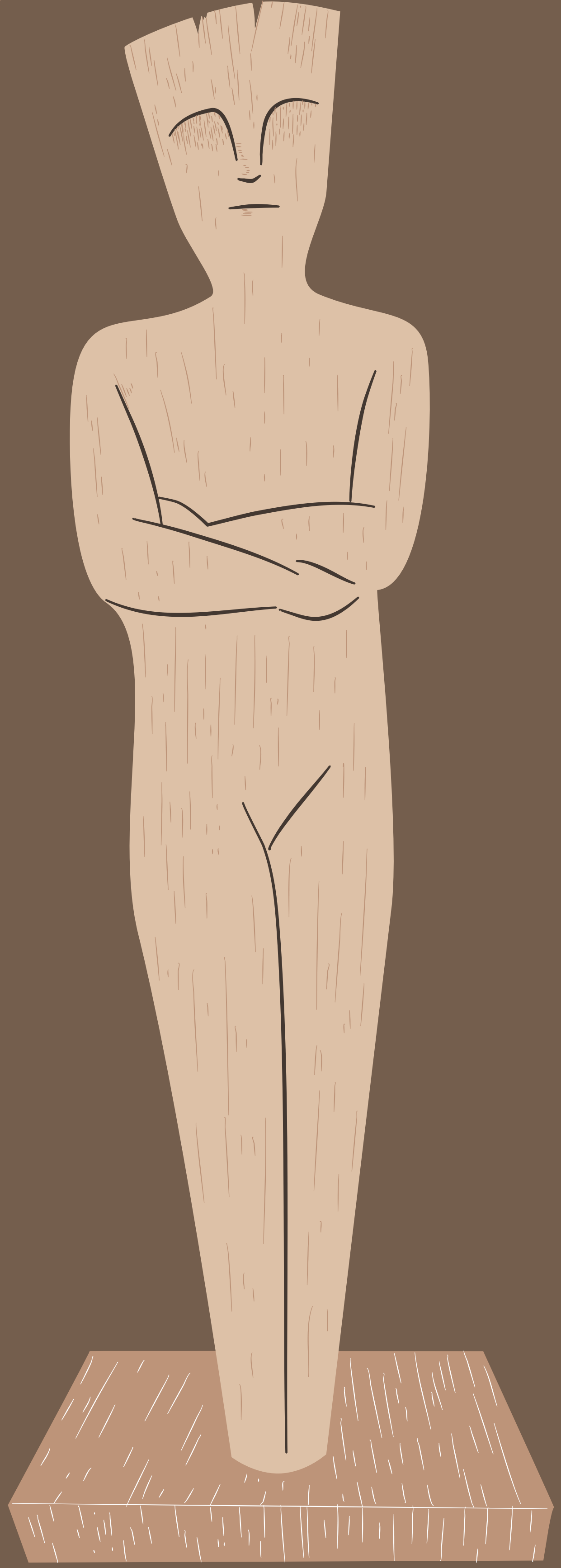
- He believed the Indus script was structured, meaningful, and shared visual and phonetic traits with scripts like Brahmi and Tamil-Brahmi.

- Yajnadevam identified repeated patterns and grammar-like rules in the script, suggesting it was a full writing system, not random symbols.

- He compiled a detailed sign catalog and proposed that some glyphs represent names of deities, agricultural terms, or professions concepts still found in Dravidian languages.

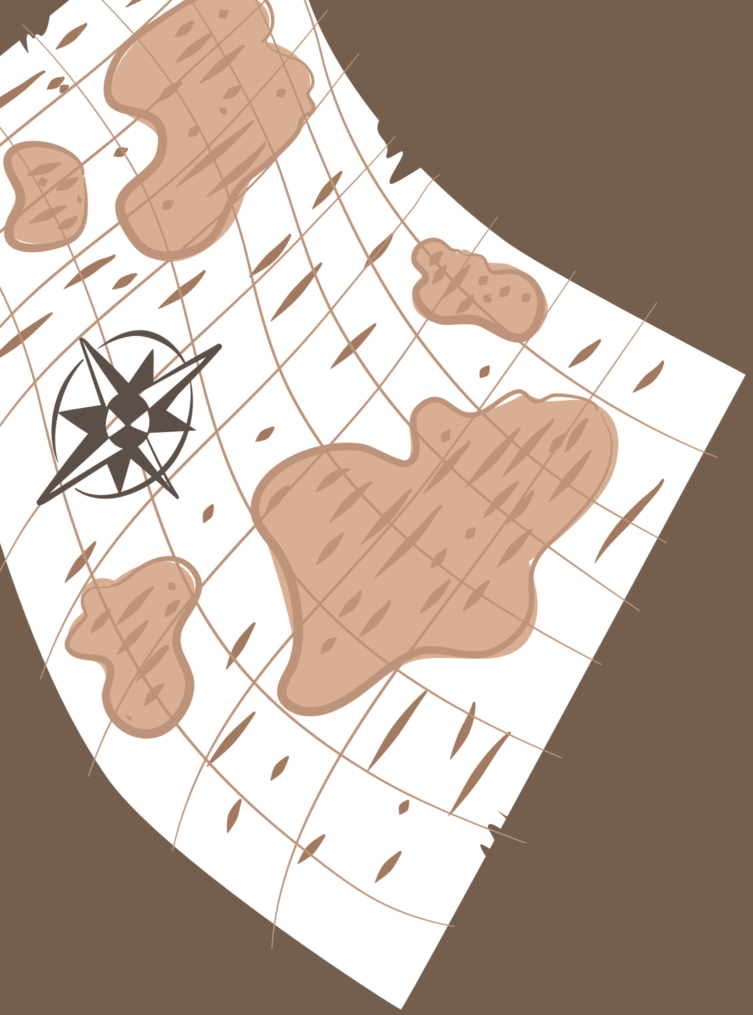
- His research motivates modern, data-driven methods like ours to check script similarities using visual analysis and machine learning.





# DATASET PREPERATION

- Collected glyph images for ancient scripts like Brahmi, Devanagari, Grantha, Kharosthi, Odia, Brahui, and IVC.
- Split each script's images into train and test folders to prepare for model training and evaluation.
- Each image was segmented into 64x64 patches to focus on individual character shapes (glyphs).
- Prepared clean and consistent datasets stored in split_train and split_test folders for feature extraction.



# DATA PIPELINE OVERVIEW

→ Raw Image
→ Glyph Segmentation
→ Resize
→ CNN Feature Extraction
→ Feature Vectors
→ SVM Classification
→ Similarity Comparison

# CHARACTER RECOGNITION

- Used Convolutional Neural Networks (CNNs) to extract visual features from each 64x64 glyph image, capturing shape, stroke, and pattern information.
- Feature vectors were generated for each character and stored for both train and test sets of every script.
- To compare glyphs across languages, correlation and covariance techniques were used to measure the similarity between feature vectors.
- A higher correlation value indicates closer resemblance in glyph structure helping identify visual and structural links between the Indus script and other ancient languages.

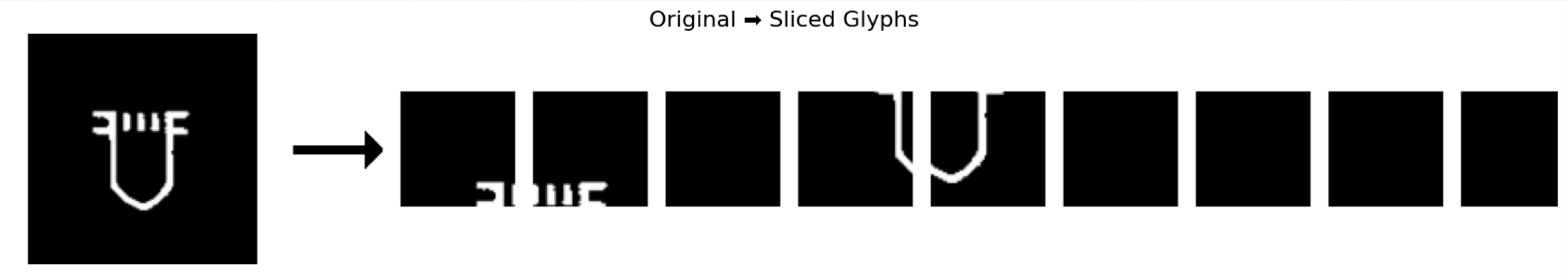
$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

pearson correlation formula

- X=(X1,X2,...,Xn): Feature vector of a glyph from an Indic script (e.g., Brahmi, Grantha, etc.)
- Y=(Y1,Y2,...,Yn): Feature vector of a glyph from the Indus Valley script

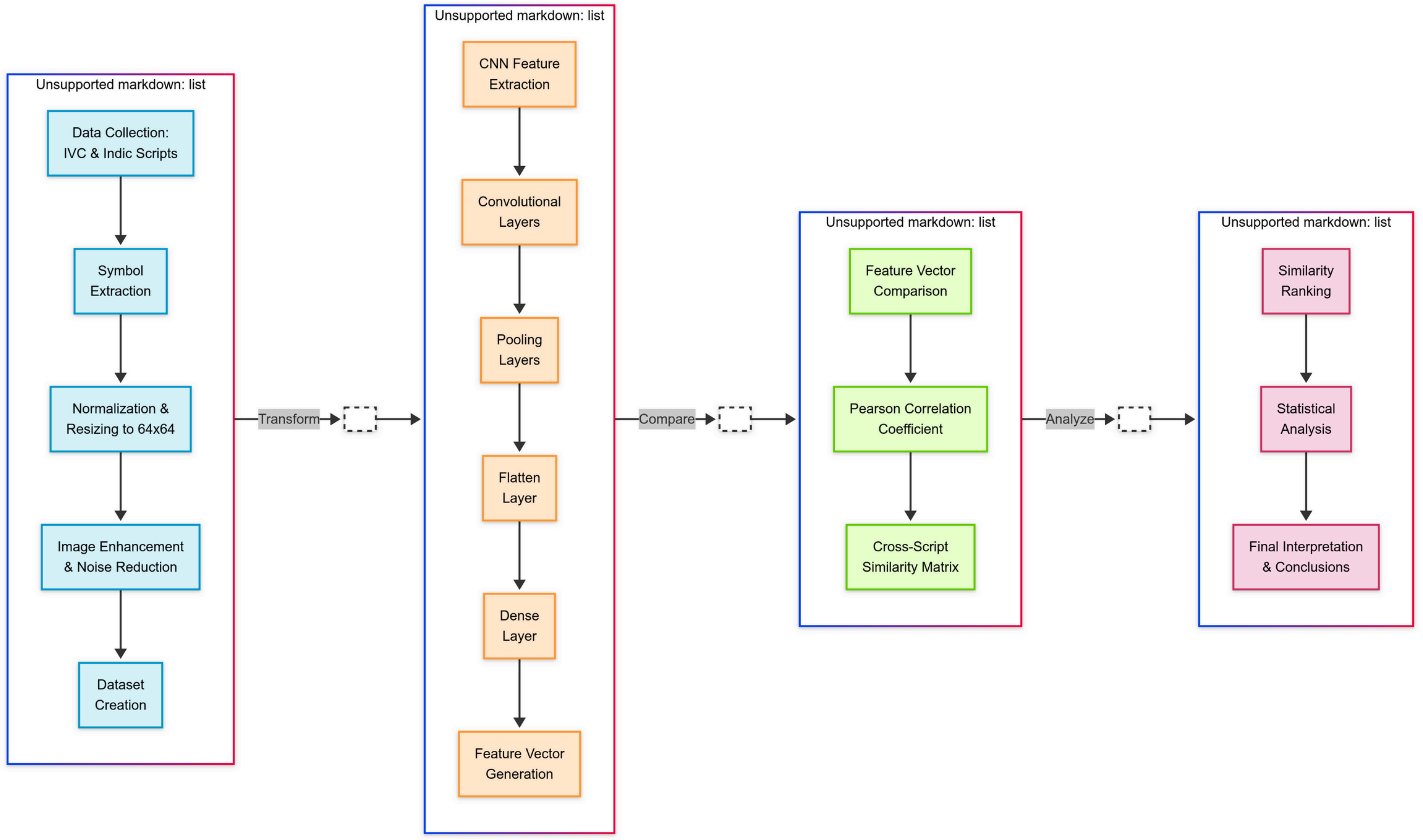
# FEATURE EXTRACTION

- Used Convolutional Neural Networks (CNNs) to extract detailed visual features from each 64x64 glyph image.
- CNN automatically learns patterns such as edges, curves, junctions, and stroke styles that define a character's shape.
- Extracted features were converted into high-dimensional vectors, preserving the uniqueness of each glyph.
- These vectors provide a standardized way to numerically compare glyphs across different scripts.
- Feature vectors from all scripts were stored in structured CSV files to allow consistent training, testing, and similarity analysis later.




IVC 64x64 glyph dimension

# WORKFLOW


Unsupported markdown: list
Data Collection: IVC & Indic Scripts
Symbol Extraction
Normalization & Resizing to 64x64
Image Enhancement & Noise Reduction
Dataset Creation
Transform
Unsupported markdown: list
CNN Feature Extraction
Convolutional Layers
Pooling Layers
Flatten Layer
Dense Layer
Feature Vector Generation
Compare
Unsupported markdown: list
Feature Vector Comparison
Pearson Correlation Coefficient
Cross-Script Similarity Matrix
Analyze
Unsupported markdown: list
Similarity Ranking
Statistical Analysis
Final Interpretation & Conclusions

# TRAINING & EVALUATION

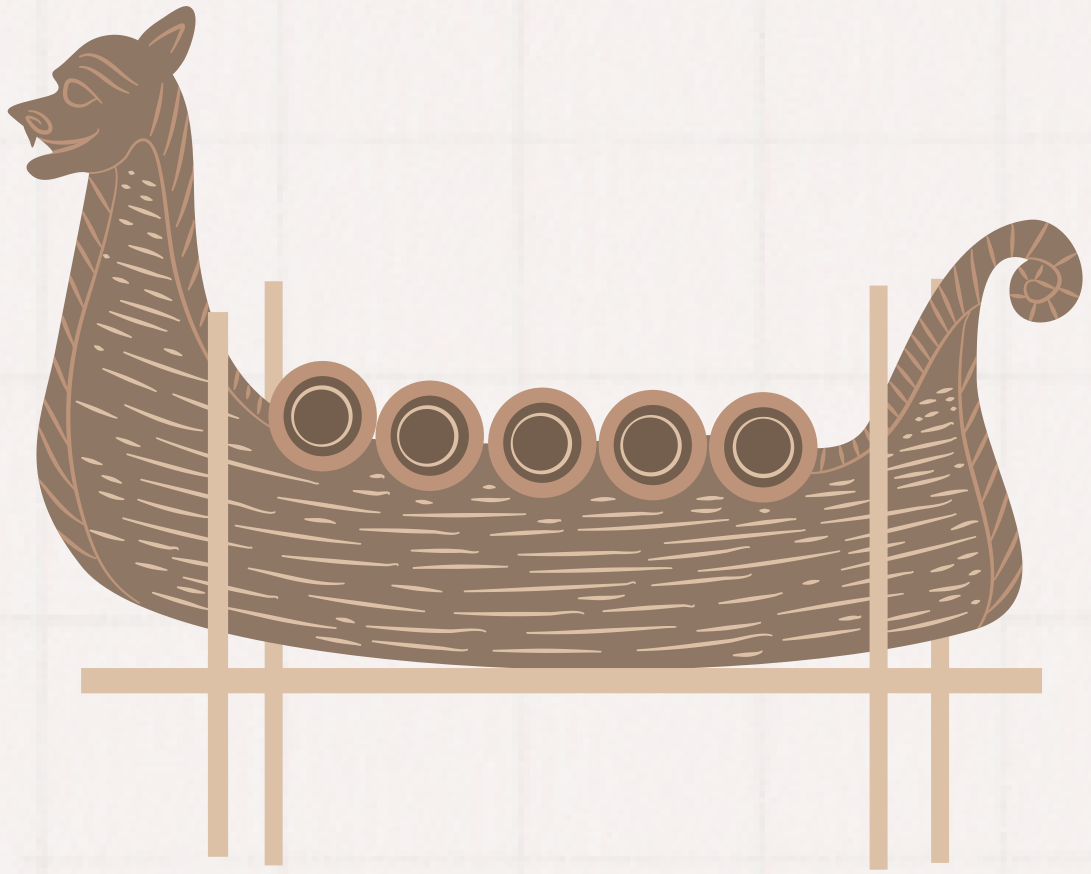
- A Support Vector Machine (SVM) classifier was trained using the extracted feature vectors from the split_train folders of each script.
- The classifier was then tested on feature vectors from the split_test folders to evaluate its ability to recognize characters based on glyph similarity.
- Evaluation included calculating the correlation between feature vectors of each script and those of the Indus Valley Civilization (IVC) script.
- The correlation values indicate the degree of visual similarity between scripts. For example, Tamil (0.45), Grantha (0.43), Brahmi (0.42), and Kharosthi (0.40) showed higher correlation with IVC.
- Among all, Brahmi script emerged as the most visually similar to the IVC script based on consistent high correlation and glyph pattern resemblance.

# RESULTS

- Grantha (0.43), Tamil-Brahmi (0.45), and Brahmi (0.42) showed the highest correlation with the Indus Valley script, suggesting strong visual similarity in glyph structures.

- These scripts are historically interconnected, with Brahmi considered a root script influencing both Grantha and Tamil.

- The high correlation supports the theory that these Indic scripts may have evolved from or were influenced by the IVC script.



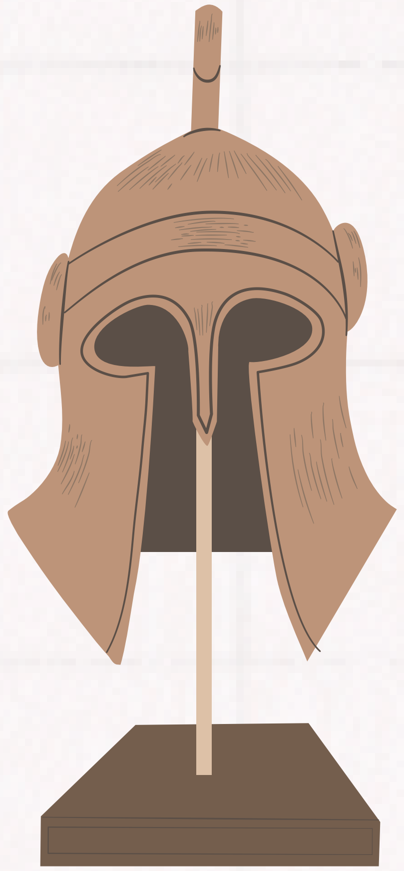
# CONCLUSION & FUTURE WORK

Conclusion:

- We used machine learning to study how similar ancient Indian scripts are to the Indus Valley script by comparing character shapes (glyphs).
- Brahmi, Grantha, and Tamil showed the highest similarity with the Indus script.
- This supports the idea that these scripts may have developed from or been influenced by the Indus script.

Future works:

- Improve the image quality and glyph segmentation for better accuracy.
- Try more advanced models like Siamese Networks for deeper similarity analysis.
- Expand the dataset with more scripts and historical variations.



# THANK YOU FOR LISTENING