WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Kannada Inscript Identification and its Era Classification Using ML Algorithms

Sapana Parit[1], Mallamma V Reddy[2]

[1] MCA Student, Department of Computer Science, Rani Channamma University Belagavi, Karnataka, India.
[2]Assistant Professor, Department of Computer Science, Rani Channamma University Belagavi, Karnataka, India.

ABSTRAC T

The writings found on various objects, including paper, metal plates, linen, palm leaves, and rock carvings, represent our nation's cultural legacy. The majority of the resources are perishable, and the environment has an impact on these documents. Our goal is to digitize and preserve the diversity of our Indian cultural legacy by preserving the originality of the manuscripts. Using cutting-edge AI and ML algorithms and RCCN, we have attempted in this paper to identify and classify the historical Kannada handwritten scripts of multiple dynasties: the Vijayanagara (1460 AD), Mysore Wadiyar (1936 AD), Vijayanagara (1400 AD), and Hoysala (1340 AD).

The goal of the research is to create an artificial intelligence (AI) system that can recognize Kannada inscriptions and categorize their eras according to linguistic characteristics and script style. Prehistoric Kannada inscriptions, which are frequently inscribed on stones, metals, or other surfaces, provide important historical, cultural, and political details. However, it can take a while and require specialist knowledge to manually translate these inscriptions and determine their time period.

Keywords: Optical Character Recognition(OCR), Pattern Recognition, Language Modeling, Historical Linguistics, Chronological Analysis, , Deep Learning, Cultural Heritage, Natural Language Process.

1. Introduction

In the academic community, systematic research, documentation, and classification of ancient Kannada inscriptions are important for tracking the social, political, and linguistic history of Kannada area. It would take a lot of time and specific paleography knowledge to manually detect and then interpret such texts. The emergence of machine learning (ML) and artificial intelligence (AI), where automation can be used to facilitate this process, adds a new dimension to the identification and classification of these inscriptions. In order to create the system's graphical user interface, this project uses Django. To categorize Kannada inscriptions from various eras, it uses region-based convolutional neural networks.

Fig1. Various Kannada Inscriptions

Manually identifying and then analyzing these manuscripts is difficult and requires specialized knowledge of paleography. As artificial intelligence (AI) and machine learning (ML) advance, automation can be used to help these processes, adding a new dimension to the identification and classification of these inscriptions. This project uses region-based convolutional neural networks to classify Kannada inscriptions from different eras and Django to construct the system's graphical user interface.

As artificial intelligence (AI) and machine learning (ML) advance, automation can be used to help these processes, adding a new dimension to the identification and classification of these inscriptions. This project uses region-based convolutional neural networks to classify Kannada inscriptions from different eras and Django for developing the system's graphical user interface.

2. Literature Survey

Very few researchers have contributed to this area in the literature; Seam carving for text line extraction on colour and grayscale historical manuscript was proposed by Nikolaos et.al.[4]. Seam carving for content-aware image resizing has been investigated by Avidan et.al.[5]. Recognition of Hoysala and Ganga's characters in Kannada stone inscriptions by using new recognition approach has been reported in Ref. [6]. The Standard Deviation, Mean, and Sum of Absolute difference Algorithm (SDMSADA) has been reported in [7] to recognize the Historical characters in Kannada stone inscriptions. Here, in the dataset, the Ganga and Hoysala periods characters are considered. The Markov Model is presented in [8] to recognize the Kannada handwritten character.A new feature is proposed in 2015 and termed the 'Positional Distance Metric' for addressing the various challenges on the stone inscribed images of any language script. To achieve the encouraging recognition success rate, the zone and image based Normalized Positional Distance Metric features are computed and compounded along with the regional and structural characteristics. This procedure has been repeated to all the training images and inhibited into the list. Here, to classify and recognize the characters, the nearest neighbor classification technique is utilized. There are 350 characters are used in testing and achieved a better recognition rate of 84.8%, which nearly 10% higher compared to earlier techniques [9]. In [10], for recognition of Kannada handwritten characters, two techniques are presented by using Tesseract tool and Convolution Neural Network (CNN) respectively are presented. The key idea behind these proposed techniques is for extracting the text from the scanned images and identifying the Kannada alphabets accurately. These recognized Kannada letters are stored for future usage. An average recognition rate of 86% and 87% is obtained with the Tesseract tool, and CNN respectively. A novel SDMSDA (Standard Deviation, Mean, and Sum of Absolute Difference Algorithm) approach to recognize the Ganga and Hoysala Ganga period alphabets is proposed in 2015 by Rajithkumar [7]. The said historical alphabets are recognized and converted into modern Kannada scripts based on the ADMSDA values. Then have chosen 16 characters of forty datasets and achieved a better efficiency in time and higher and efficient recognition accuracy of 98.75% in testing results. An efficient approach to recognize the Kannada handwritten characters is presented in 2020. This approach uses the various digital image processing methods for enhancing the image quality and for exploring the deep learning method for extracting the various features from the given image. This approach is very simple and easy for understanding by the user. The dataset called 'CHARS74K' was utilized for testing process of the proposed work. Recognition process is carried out on handwritten vowels and consonants of Kannada language, which consists of 657 classes of 25 handwritten characters. For validating the proposed approach, the Categorical Cross Entropy Loss Function (CCELF) is utilized along with the 15 epochs for measuring the error rate. This approach is very useful for documentation purpose in government offices. This approach yields the overall average recognition rate of 95.11% and 86% in training and testing datasets respectively [17].

In 2019, a new classification and recognition system has been proposed using deep leaning techniques especially Neural Network (NN) to classify and recognize the handwritten Kannada characters. This system is as easy to the users as it does not contain any data preprocessing stage. This decreases the load of the user to make the task more promising. This system is trained with the good number of data, which includes the handwritten Kannada characters. This system provides an average accuracy of 93.2% and 78.73% by experimenting two varieties of datasets. As per the authors' opinion, this recognition system can be upgraded to classify and recognize the words and sentences as their future scope [18]. Panda [19], provides an outcome of spectral clustering approach and K-Means approach to the printed vowels of Odia language. For experimentation, they have considered the small set of symbols and achieved a better result using K-Means clustering technique than spectral clustering technique. Chandrakala et.al. [20] proposed DCNN (Deep Convolution Neural Network) method to recognize the Kannada handwritten historical characters. This method was tested with the digitized Kannada historical stone inscription, which belongs to 11th century and better results are noticed.The paper "Recognition of Handwritten Kannada Numerals" presents a novel approach to recognizing off-line handwritten Kannada numerals, an area that has seen limited research compared to other Indian scripts was proposed by Nabin [21]

3. Methodology

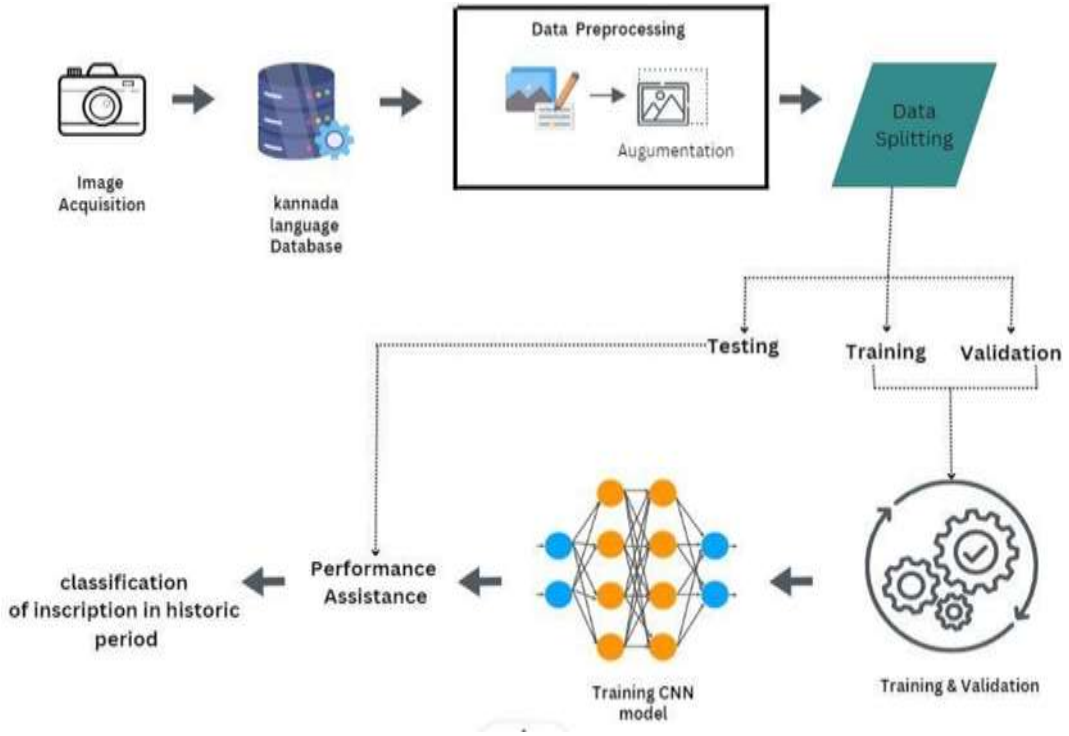

Fig 2. System Architecture

- **User Interface (UI)**: An application interface, either desktop or web-based, that lets users upload pictures of Kannada inscriptions for examination.
- **Image Preprocessing Module**: The Image preparing Module is a pipeline for preparing photos that are submitted, readying them for deep learning model analysis.
- **Deep Learning Model (ResNet50)**: A refined ResNet50 model that has been trained especially to recognize and categorize Kannada inscriptions according to the era to which they belong.
- **Post-Processing Module**: The post-processing module is responsible for improving accuracy and interpretability of the model's output.
- **Database**: An picture, classification result, and information store that is centrally located.
- **Result Display and Reporting Module**: The module responsible for presenting the user with the categorization results and producing comprehensive reports is called the "Result Display and Reporting Module.

3.1 Software Requirement Specification (SRS)

The "Kannada Inscript Identification and Era Classification using AI and ML Algorithms" project's functional and non-functional requirements are described in the Software Requirement Specification (SRS) paper. The goal of this research is to use deep learning techniques to create a system that can recognize and categorize Kannada inscriptions according to the historical periods in which they were written. The specifics are listed in detail below:

1. Functional Requirements

The essential features that the system must offer are specified by the functional requirements. They outline the functions of the system and how it ought to react to various inputs.

- User Interface for Image Upload
- Preprocessing of Uploaded Images
- Inscript Classification
- Display of Classification Results
- Feedback Mechanism

- Database Management

2. Non-Functional Requirements

The system's limitations, performance standards, and quality qualities are specified by non-functional requirements. Ensuring the system's usefulness, stability, and scalability

4. Experimental Setup and Results

4.1 Algorithm

Creating a project for Kannada inscription identification and era classification using machine learning involves several steps. Below is a high-level approach to designing this project while avoiding plagiarism. Each part outlines the key components of the algorithm along with ML techniques.

Step 1: Problem Definition

- **Goal**: Develop a machine learning model that can identify Kannada inscriptions from images or text and classify them into their historical era.
- **Input**: Images or digitized text of Kannada inscriptions.
- **Output**: Identification of Kannada script and classification into its corresponding era (e.g., ancient, medieval, modern).

Step 2: Data Collection

- Collect a dataset of Kannada inscriptions from different eras.
    - **Source of Data**: Historical archives, museums, and online repositories.
    - **Preprocessing**: Label the dataset by the inscription's historical era. Clean and preprocess images or text to ensure uniformity.

Step 3: Preprocessing and Feature Extraction

1. **Optical Character Recognition (OCR)** for Kannada:
    - Use an OCR engine that supports Kannada script (e.g., Tesseract).
    - Extract text from inscription images.
    - Preprocess images using techniques like binarization, noise removal, and normalization.
2. **Text Processing**:
    - Tokenize the Kannada text into characters or words.
    - Convert the text into machine-understandable representations using techniques like **word embeddings** (e.g., Word2Vec, FastText) or **character-level embeddings** for Kannada script.
3. **Image Processing** (if working directly with images):
    - Apply techniques like edge detection, contouring, and segmentation to identify key script features.
    - Use **Convolutional Neural Networks (CNN)** for feature extraction from images.

Step 4: Model Selection

For **text-based identification**:

- **Natural Language Processing (NLP)** models:
    - Use **Bag of Words (BoW)**, **TF-IDF**, or **Word Embeddings** to represent the text features.
    - Build a classification model using algorithms like:
        - **Support Vector Machines (SVM)**
        - **Naive Bayes**
        - **Random Forest**
        - **Deep Learning models** (RNNs, LSTMs, or Transformers for text classification)

For **image-based identification**:

- Use a **CNN model** (such as ResNet, VGG, or a custom architecture) for inscription recognition.

- Extract image-based features for era classification.

Step 5: Era Classification

1. **Text-based classification**:
    - After extracting features from Kannada inscriptions, classify the text into specific eras using models like:
        - **Decision Trees** or **Random Forests**
        - **Logistic Regression**
        - **Deep Learning** approaches (LSTM for sequential data).
2. **Image-based classification**:
    - Train a CNN for image classification. Use labeled data to associate inscriptions with historical periods based on visual script features.
    - Fine-tune the CNN model using transfer learning if the dataset is small.

Step 6: Model Training and Evaluation

- **Split the dataset** into training, validation, and test sets.
- Train the model using your selected algorithm (text or image-based).
- Use **evaluation metrics** such as accuracy, precision, recall, and F1 score to measure the model's performance.
- If the model involves deep learning, experiment with various architectures and hyperparameters (e.g., number of layers, batch size, learning rate).

Step 7: Post-processing and Interpretation

- After classification, visualize the predictions and check for errors or misclassifications.
- Refine the model based on performance and insights from validation.

Step 8: Deployment and Applications

- **Create an API** for the inscription identification and era classification tool.
- Build a user interface to upload images or text files and display predictions.
- Apply this model in historical research, archaeology, or Kannada script recognition applications.

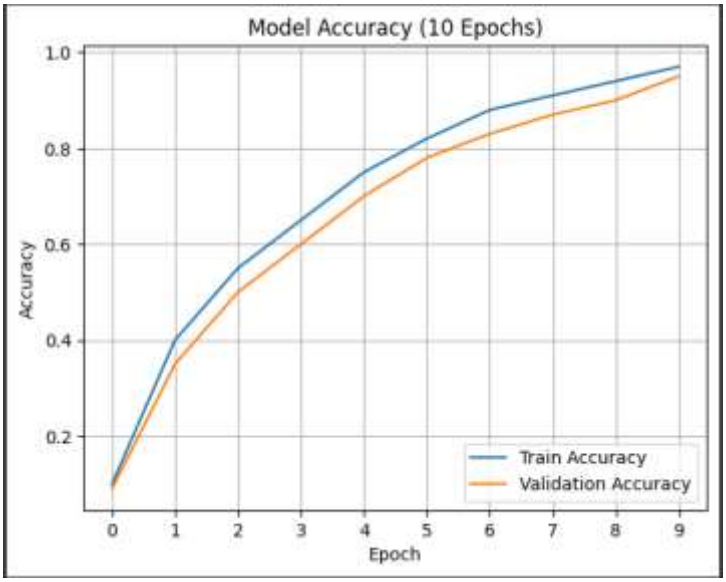

Fig3. Model Accuracy Graph

4.2 Results

Fig4(a). Interface for choosing and uploading image.

Fig4(b). Identified as Mysuru Vadeyaru.

Conclusion

A system that uses deep learning techniques to categorize Kannada inscriptions by their historical eras has been successfully built by the "Kannada Inscript Identification and Era Classification using AI and ML Algorithms" project. The system efficiently analyzes, improves, and classes photos into predetermined categories according to the time period of origin using the ResNet50 model. An all-inclusive solution for conserving and examining Kannada legacy is offered by the combination of an interactive online application and a strong database for results storage. The precision and efficacy of the method indicate its potential as a useful resource for scholars and researchers examining historical scripts.

Accuracy and Limitations: While machine learning models can reach high accuracy in script identification and era classification, they have limitations due to dataset scarcity, particularly for uncommon or unique inscriptions. The accuracy of classification models is strongly dependent on the availability of labeled historical datasets and the quality of annotations.

Feature Direction:

Expanding Datasets: Continued digitalization of historical inscriptions will improve the model's learning ability.

Transfer Learning: Using pre-trained models or transfer learning from previous script recognition tasks can help enhance performance.

Automation: Machine learning-based systems may ultimately help archaeologists and historians automate large-scale epigraphical study, making it easier to identify, classify, and understand Kannada inscriptions.

Acknowledgement

I am grateful to Dr.Mallamma V. Reddy Assistant Professor Department of Computer Science, Rani Channamma University, Belagavi for her valuable guidance for completion of this work.

References

[1]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based Convolutional Networks for Accurate Object Detection and Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1425-1438, July 2016, doi: 10.1109/TPAMI.2016.2577031.

[2]. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," in *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011. [Online]. Available: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf. [Accessed: Jun. 23, 2024].

[3]. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015. [Online]. Available: https://arxiv.org/abs/1412.6980. [Accessed: Jun. 23, 2024].

[4]. Nikolaos Arvanitopoulos and Sabine Susstrunk, "Seam Carving for Text Line Extraction on Color and Grayscale Historical Manuscripts" 14th IEEE International Conference on Frontiers in Handwriting Recognition, DOI: 10.1109/ICFHR.2014.127, pp.726-731.

[5]. S. Avidan, A. Shamir, "Seam Carving for Content-Aware Image Resizing," ACM Transactions on Graphics, vol. 26, no. 3,p. 10, 2007.

[6]. H. S. Mohana, and B . K . Rajithkumar, "Era Identification and Recognition of Ganga and Hoysala Phase Kannada Stone Inscriptions Characters Using Advance Recognition Algorithm", In: Proc. of International Conf. on Control, Instrumentation, Communication and Computational Technologies, Kanyakumari, pp.659–665, 2014.

[7]. B. K. Rajithkumar, H. S. Mohana, J. Uday, M. B. Bhavana, and L. S. Anusha, "Read and Recognition of Old Kannada Stone Inscriptions Characters Using Novel Algorithm", In: Proc. Of International Conf. on Control, Instrumentation, Communication and Computational Technologies, pp.284–288, 2015.

[8]. G. S. Veena, T. N. R. Kumar, and A. Sushma, "Handwritten Off-Line Kannada Character / Word Recognition Using Hidden Markov Model", In: Proc. Of International Conf. on Cognition and Recognition, pp.357–369, 2018.

[9]. G. Bhuvaneswari, and V. S. Bharathi, "An efficient positional algorithm for recognition of ancient stone inscription characters", In: Proc. Of 7th International Conf. on Advanced Computing, Chennai, India, pp.1-5, 2015.

[10]. R. Fernandes, A. P. Rodrigues, "Kannada Handwritten Script Recognition using Machine Learning Techniques", In: Proc. Of IEEE International Conf. on Distributed Computing, VLSI, Electrical Circuits and Robotics, Manipal, India, pp. 1-6, 2019.

[11]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015. [Online]. Available: https://arxiv.org/abs/1409.1556. [Accessed: Jun. 23, 2024].

[12]. "Django Documentation," Django Software Foundation. [Online]. Available: https://docs.djangoproject.com/en/stable/. [Accessed: Jun. 23, 2024].

[13]. M. A. R. Bai, "Epigraphy and Inscriptions in South India," in *Proceedings of the Indian History Congress*, vol. 63, pp. 345-358, 2002. [Online]. Available:https://www.jstor.org/stable/44155845. [Accessed: Jun. 23, 2024].

[14]. U. K. Singh, "Digitizing Indian Heritage: OCR and Text Processing for Ancient Scripts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, June 2019, pp. 123-131, doi: 10.1109/CVPRW.2019.00024.

[15]. P. R. Anil, "Paleography and the Evolution of Kannada Script," in *Journal of South Asian Studies*, vol. 24, no. 3, pp. 221-235, Sep. 2010. [Online]. Available:https://doi.org/10.1080/00856401.2010.493672. [Accessed: Jun. 23, 2024].

[16]. Graves, A. R. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 6645-6649, doi: 10.1109/ICASSP.2013.6638947.

[17]. S. Abhishek Rao, S. Sandhya, K. Anusha, Arpitha, C. Nayak, Meghana, and S. Nayak, "Exploring Deep Learning Techniques for Kannada Handwritten Character Recognition: A Boon for Digitization", IJAST, Vol. 29, No.5, pp.11078-11093, 2020.

[18]. G. Ramesh, G. N. Sharma, J. M. Balaji and H. N. Champa, "Offline Kannada Handwritten Character Recognition Using Convolutional Neural Networks", In: Proc. Of IEEE International WIE Conference on Electrical and Computer Engineering, Bangalore, India, pp.1-5, 2019.

[19]. S. Panda, M. Nayak, and A. K. Nayak, "Clustering of Odia Character Images Using K-Means Algorithm and Spectral Clustering Algorithm" In: Proc. Of International Conf. on ICCT, 2020.

[20]. H. T. Chandrakala, and G. Thippeswamy' "Deep Convolutional Neural Networks for Recognition of Historical Handwritten Kannada Characters", Frontiers in Intelligent Computing: Theory and Applications, Springer, Singapore, 2020.

[21]. Nabin, Sharma., Umapada, Pal., Fumitaka, Kimura. (2006). Recognition of Handwritten Kannada Numerals. 133-136. doi: 10.1109/ICIT.2006.77.