IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automatic Recognition Of Handwritten Kannada Districts Names With Machine Learning Models

Dr. K M Rajesh[1] Asst. Professor and Hod, Department of Computer Science, JSS College for women, Chamarajanagara

Manasa P S[2] Asst. Professor, Department of Computer Science, JSS College for women, Chamarajanagara

Abstract

In this paper, we present a machine learning models for handwritten Kannada language district name recognition. Kannada, is a prominent Indian language, has a unique script with complex characters, making handwritten recognition a challenging task. The goal of this research is to develop an efficient system that can accurately recognize district names in Kannada from handwritten text. The proposed method involves preprocessing of handwritten input images, feature extraction using deep learning techniques, and classification using machine learning algorithms such as Convolutional Neural Networks (CNNs). A dataset containing a large variety of handwritten Kannada district names is utilized for training and testing the model. We evaluate the performance of various machine learning models, comparing their accuracy and processing speed. The results demonstrate the effectiveness of deep learning models in achieving high recognition accuracy, even with diverse and noisy handwritten inputs. This system holds potential for applications in digitizing official documents, automating postal services, and improving language-based AI systems for Kannada. The study also explores future directions for improving recognition accuracy, including the integration of hybrid models and dataset expansion.

Keywords

Handwritten Text Recognition, Kannada Language, Machine Learning, Convolutional Neural Networks (CNN), District Name Recognition, Feature Extraction, Deep Learning, Optical Character Recognition (OCR), Image Processing, Artificial Intelligence (AI), Pattern Recognition, Language Digitization, Handwritten Kannada Dataset.

I. INTRODUCTION:

Handwritten text recognition is a challenging task, especially in languages with complex scripts like Kannada. Kannada is spoken in the Indian state of Karnataka and has a unique script with over 60 characters, making it difficult for machines to recognize handwritten content accurately. Many official documents, such as postal forms, contain handwritten district names in Kannada, and automating the recognition of these names would be highly beneficial.

Current systems mainly focus on recognizing printed text, and there is limited research on recognizing handwritten text in Kannada. This paper presents a machine learning- to recognize handwritten Kannada district names. We use deep learning techniques, particularly Convolutional Neural Networks (CNNs), to build a system that can identify district names from handwritten text. Our model is trained on a large dataset of handwritten Kannada district names and aims to handle various handwriting styles and noise in the input.

This work can be applied to areas like digitizing government documents, improving postal services, and developing AI tools for Kannada language processing. The paper is organized as follows: Section 2 covers related work, Section 3 explains our methodology, Section 4 presents the results, and Section 5 discusses future improvements.

II. PROBLEM STATEMENT

Given a handwritten Kannada language district names, the system needs to predict the type of the district. In other words if we can write the district name “Chamarajanagar” the system predict the district name that is truly “Chamarajanagar” or the input character is nearer to “ Chamarajanagar” or something else. The purpose of this project is to take the Hand Written Kannada language district names as an input, process it, train the system effectively by using the algorithm to recognize the pattern.

III. LITERATURE SURVEY

1. Title: Offline Handwritten recognition of Malayalam District Name – A Holistic Approach
Author: Jino PJ, Kannan Balakrishnan
Year: 2017
Holistic method of handwritten word recognition is implemented with the help of three popular machine learning classifiers like Neural Network, SVM, and Random Forest. SVM with RBF kernel provide 97% Result with PCA as the dimensionality reduction approach. Lexicon was less mainly because the total number of districts in Kerala is 14. The work can be extended to more lexicon size. Also we can consider handcrafted features along with machine extracted features for the recognition.

2. Title: Offline Kannada Handwritten Word Recognition Using Locality Preserving Projection (LPP) for Feature Extraction
Author: M.S. Patel, Rohith Kumar, S.C. Linga Reddy
Year: 2015
Offline handwritten word recognition is the most challenging task in the field of image processing. Kannada language having its cursive nature and large number of alphabets. Lack of availability of standard data set challenges the researchers to do

the research in Kannada language. We have used Locality Preserving Projections (LPP) to reduce the dimensionality and to extract the features .Support Vector Machines (SVM) is provided and K-means is used to recognize the word images based on the extracted features. Experimental result shows the good accuracy rate towards Kannada handwritten words. In future, we explore different variants of subspace learning method for better representation task.

3. Title: Offline handwritten Kannada word recognition

Author: Krupashankari.S Sandyal, M.S.Patel

Year: 2014

The database is created by collecting data from 60 different writers so that recognition engine could be trained with different styles of handwriting. As we were specific to postal mail application we collected writings of all the 30 district names of Karnataka from each writer. So a total of 1200 words dataset was created. The proposed method was trained with a dataset of 250 bitmap images digitized at 300 dpi and the recognition engine is tested for the accuracy. Euclidean distance classifier is used to classify the test images and from experiments done 92% of recognition rate is obtained.

4. Title: Offline Handwritten Word Recognition: A Survey

Author: M S Patel, Rohith Kumar

Year: 2015

Offline handwritten word recognition is one of the interesting fields of research in the Image processing. Though lot of work has been done still it has got many opportunities to do the work in this field. Reasons for this are, in the most of languages lack of availability of standard datasets accuracy rate obtained. By using good combinations of feature extraction methods and classifiers it is possible to achieve the good results. In our survey paper different methods used for preprocessing, feature extraction, classification are discussed. Survey has done on English, Arabic, Hindi, Kannada languages. We believe that this survey will helpful for researchers in this field.

6. Title: Handwritten Character Recognition Using Feed-Forwarded

Author: Nilay Karade, Dr. Manu Prathap Singh, Dr. Pradeep K. Butey

Year: 2015

Handwritten Character recognition has been vigorous and tough task in the field of pattern recognition. Considering its application to various fields, a lot of work is done and is being continuing to improve the results through various methods. In this paper we have proposed a system for individual handwritten character recognition using multilayer feed-forwarded neural networks. For the experimental purpose we have taken 15 samples of lower and upper case handwritten English alphabets in scanned image format i.e. 780 different handwritten character samples. There are two methods of feature extraction are used to construct the pattern vectors for training set. This training set is presented to the six different feed-forwarded neural networks namely newff, newfit, newpr, newgrnn, newrb, and newrbe. The test pattern set is used to evaluate the performance of these neural network models. The number of hidden layer, number of neurons of neurons in hidden layer, validation checks and

gradient factors of the neural networks models are taken into consideration during the training.

IV. EXISTING SYSTEM

Major challenge for analytic recognition is the need for a proper segmentation algorithm. Even some times human beings may not be able to segment characters properly; in that case they will recognize it from the context or shape of the word. The proper selection of the feature purely depends on the language domain. In modern research we can find a lot of methods implementing machine extracted features rather than handcrafted features.

Disadvantages

• Not 100% accurate, there are likely to be some mistakes made during the process

• Take more time to execute

V. PROPOSED SYSTEM

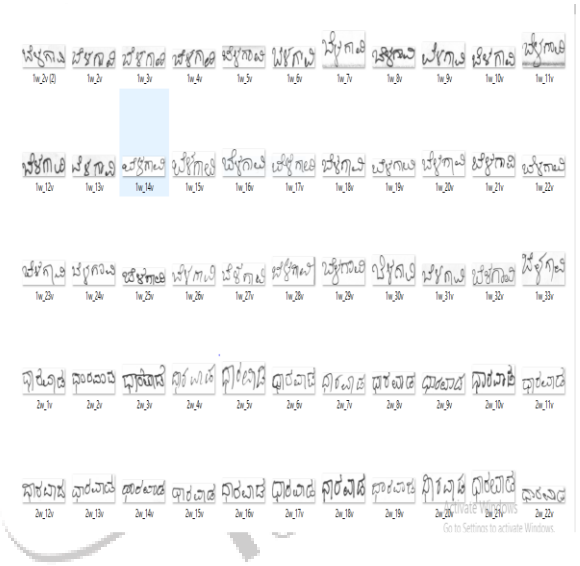
The initial step to be carried out is to place the dataset, which can be effectively done through programming interface. User uploads a particular image of any district names which we wants to recognize. The image will be processed by the system. On running the system code the output is generated that shows which is the district name uploaded by the user and also displays the accuracy rate predicted by the model. On uploading image with different resolutions other than the one mentioned in the code. The output generated shows error, and displays an error message to the user. There are two ways to provide input to the system. The user can either upload the image of the district name we wants to detect or the data from the dataset. The input images are pre-processed. Using the different classifiers the recognized digit accuracy is compared and the results obtained are displayed along with the accuracy.

Advantages

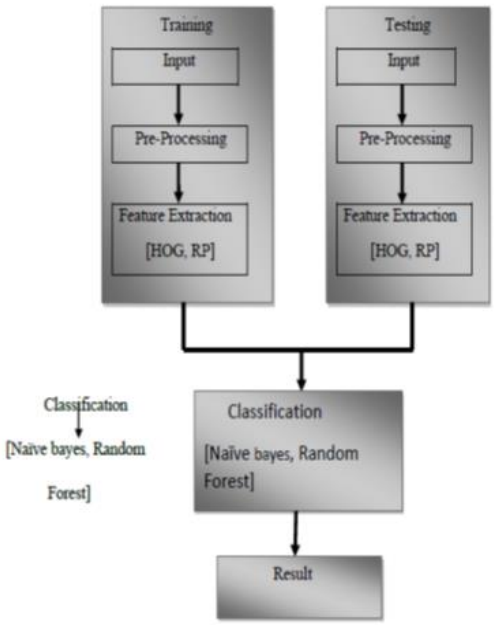
- Large quantity of text can be input quickly
- A paper based form can be turned into an electronic one
- It can read handwriting
- The process is more accurate.

VI. DATASET REVIEW

The dataset consists of 2 classes
• Training samples
• Testing samples



Total samples of dataset: 1,569.

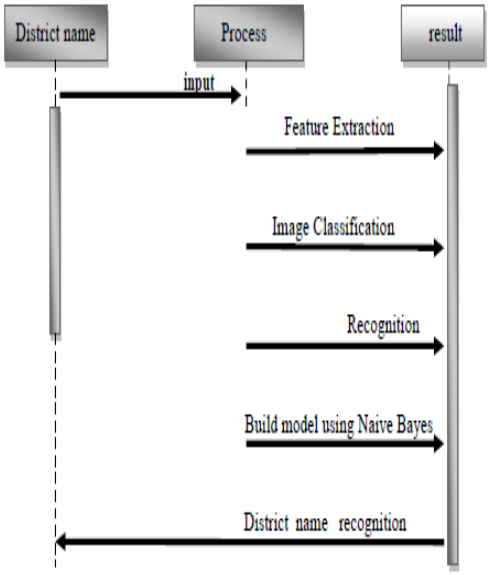


Testing samples: 311

VII. SCOPE

The system Design (SD) describes how the functional and non-functional requirements gathered in the requirement phase, preliminary user-oriented functional design transform into more technical system specifications from which the system is built. This phase describes the design goals and considerations, provides a high-level overview of the system architecture, and describes the data design associated with the human-machine interface and operational scenarios.

VIII. SYSTEM ARCHITECTURE



IX. CONCLUSION AND FUTURE SCOPE

Handwritten word recognition is the most challenging task in the field of image processing, Kannada language having its cursive nature and large number of alphabets Lack of availability of standard data set challenges the researches to do the research in Kannada language.

Our future goal is to improve the accuracy and also to extracting the new features

Result:

	Random Forest	Naive Bayes
80 - 20	83.01	87.17
20 - 80	47.67	71.82
30 - 70	61.92	81.10
70 - 30	82.05	86.53
50 - 50	76.50	83.56
90 - 10	84.61	87.82

Processing:

X. REFERENCES

[1] M.S. Patel1, Rohith Kumar2, S.C. Linga Reddy, "Offline Kannada Handwritten Word Recognition Using Locality Preserving Projection for Feature Extraction", vol. 4, Issue 7, july 2015

[2] Jino P J, Kannan Balakrishnan, "Offline Handwritten Recognition of Malayalam District Name- A Holistic Approach".

[3] Krupashankari.S.Sandyal1, M.S.Patel2, "Offline Handwritten Kannada Word Recognition".

[4] Sarika Hegde 1, K. K. Achary2, Surendra3, "statistical analysis of features and classification of alphasyllabary sounds in Kannada language".

[6] Ravikumar M[1], Sampathkumar S[1] and Shivakumar G[1], "Kannada Handwritten Answer Recognition using Machine Learning Approach".

[5] Rituraj Kunwar1, Shashikiran K2, A. G. Ramakrishna, "Online handwritten Kannada word recognizer with unrestricted vocabulary".