

Bengali Handwritten Characters Recognition

Bornomalar Bondhu
Dipankar Dey
Saikat Kumar Ghosh

February 25, 2024

Abstract

This project is going to explore the feasibility of using traditional supervised machine learning techniques for Bengali Handwritten Character Recognition (BHCR). We will use a diverse dataset of handwritten Bengali characters, classical algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests etc. which will be trained and evaluated. Through rigorous performance assessment using standard metrics, including accuracy, precision, the effectiveness of these classical methods will be demonstrated. Feature selection and hyperparameter tuning might be used to further enhance model performance. At the end we will assess and compare the accuracy given by classical Machine learning techniques with popular deep learning techniques .

1 Introduction

The objective of this project is to leverage machine learning techniques to classify handwritten Bengali characters. The Bengali script comprises 11 vowels and 39 consonants, totaling 50 basic characters. By employing machine learning algorithms, we aim to develop a robust classification system capable of accurately identifying and categorizing handwritten Bengali characters into their respective vowels and consonants categories. The complexity of handwritten character recognition, particularly in Bengali, poses significant challenges due to variations in size, shape, and individual writing styles. These challenges are increased by similarities in character shapes, and variations in strokes. Bengali, as one of the world's most spoken languages with rich cultural heritage, demands attention in automatic character recognition. Addressing these challenges through machine learning holds promise for enhancing recognition accuracy and advancing linguistic technology. Here we will be using the classical machine learning models to achieve our objective and will try to get as close as possible with the accuracy given by popular deep learning models by incorporating more complex datasets in our study.

2 Methodology

1.Literature Review: Conduct a comprehensive review of existing literature to understand the current state-of-the-art, relevant theories, and methodologies in the field.

2.Data Collection: Gather relevant data from various sources, ensuring it aligns with the project objectives and research questions.

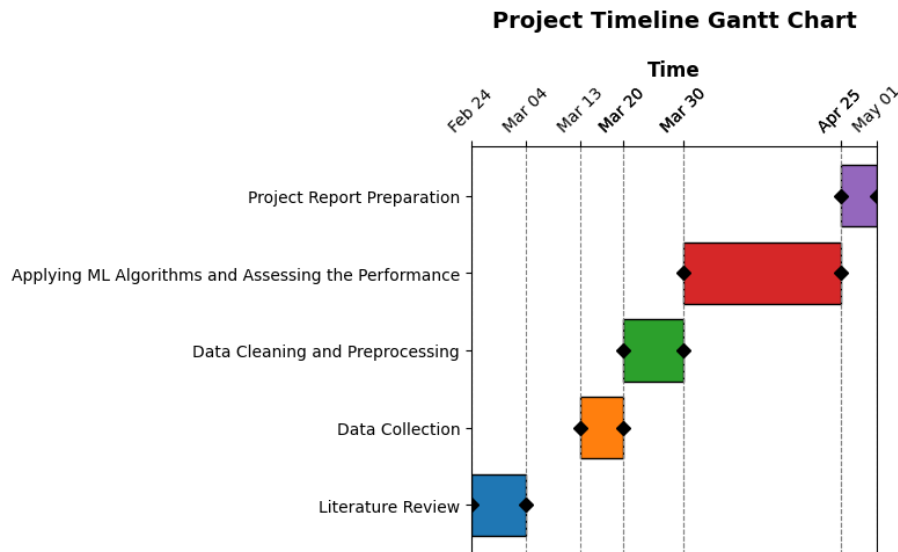
3.Data Cleaning and Preprocessing: Cleanse the collected data to remove errors, inconsistencies, and irrelevant information. Preprocess the data by transforming, scaling, or encoding it as necessary for analysis.

4.Applying ML Algorithms: Select appropriate machine learning algorithms based on the nature of the problem and the data characteristics. Train the models using the pre-processed data.

5.Assessment of Performance: Evaluate the performance of the trained models using appropriate metrics such as accuracy, precision, recall, or F1-score. Fine-tune the models if necessary to improve performance.

6.Preparation of Project Report: Summarize the entire project including objectives, methodology, findings, and conclusions in a comprehensive project report. Clearly present the insights gained, challenges encountered, and recommendations for future work within the specified format and guidelines.

3 Work plan and time line



4 Work plan division in our group

Literature Review: Will be done by both of us individually.

Data Collection: Dipankar Dey

Data Cleaning and Preprocessing: Saikat Kumar Ghosh

Application, Development and Performance Evaluation of ML Algorithms: Will be done by both of us.

5 Reference

References

- [1] Rabby A.S.A., Haque S., Abujar S. Islam M.S., and Hossain S.A. Ekush A multi-purpose and multitype comprehensive database for online off-line bangla handwritten characters. <https://rabby.dev/ekush/#external>, 2019.
- [2] Ujjwal Bhattacharya, M. Shridhar, Swapan Kumar Parui, Bidyut Baran Chaudhuri, and P. K. Sen. Off-line handwritten database for bangla basic characterst. <https://www.isical.ac.in/~ujjwal/download/Banglabasiccharacter.html>, 2012.

- [3] Ujjwal Bhattacharya, M. Shridhar, Swapan Kumar Parui, Bidyut Baran Chaudhuri, and P. K. Sen. Offline recognition of handwritten bangla characters: An efficient two-stage approach. *Pattern Analysis and Applications*, 15(4):35, 2012.
- [4] Tapotosh Ghosh, Md. Min ha-zul Abedin, Shayer Mahmud Chowdhury, and Mohammad Abu Yousuf. A comprehensive review on recognition techniques for bangla handwritten characters. IEEE, 2019.
- [5] Nabeel Mohammed, Sifat Momen, Anowarul Abedin, Mithun Biswas, Rafiqul Islam, Gautam Shom, and Md. Shopon. Bangla lekha isolated dataset. <https://data.mendeley.com/datasets/hf6sf8zrkc/2>, 24 February 2017.
- [6] AKM Shahariar Azad Rabby, Sadeka Haque, Shammi Akther Shahinoor, Sheikh Abujar, and Syed Akhter Hossain. A universal way to collect and process handwritten data for any language. *Procedia Computer Science*, 143(23):502–509, 2018.
- [7] Moshiur Rahman, Iqbal Mahmud, Md Palash Uddin, and Masud Ibn Afjal. Handwritten bengali character recognition through geometry based feature extraction. *Journal of Theoretical and Applied Information Technology*, 97(23):3570–3582, 2019.