A

Literature Survey on

Character Recognition Using Ensemble Technique

Submitted to

For the partial fulfillment of Award of the degree

MASTER

IN

COMPUTER APPLICATION

By

Prashant Singh

(MCAN1CA22019)

Under the Supervision of

Dr. Vani Agarwal, Assoc. Professor,

Department of CSA

ITM University

Contents

A Literature Survey for Character Recognition Using Ensemble Technique

Prashant Singh (MCAN1CA22019)

ITM University, Gwalior

Abstract of the Project

This research is dedicated to the enhancement of handwritten letter recognition through the innovative application of ensemble techniques, with a primary focus on improving the accuracy of deciphering handwritten characters, particularly emphasizing the Marathi script. The significance of this work extends beyond mere character recognition, as it carries substantial implications for the field of document analysis. The research endeavors to make document processing more efficient by enabling the extraction of information from handwritten documents with higher precision, which, in turn, benefits industries reliant on automated document processing. Furthermore, the research seeks to strengthen character recognition systems, including optical character recognition and handwriting transcription, by implementing diverse pre-processing methods. These techniques, such as image augmentation, denoising, and normalization, collectively contribute to improving the quality and consistency of input data, addressing the inherent challenges of handwritten characters that can vary greatly in style, legibility, and quality. Additionally, meticulous feature selection is employed to identify and retain the most relevant attributes, effectively reducing data dimensionality and enhancing Deep Learning model performance. The core methodology of this research revolves around the ensemble of Deep Learning algorithms, where traditional algorithms are combined to maximize their collective potential, resulting in superior recognition accuracy. To ensure the robustness and adaptability of the ensemble, the study employs a K-Fold cross-validation strategy, rigorously evaluating its performance across diverse data subsets. In conclusion, the study's unwavering focus on handwritten letter recognition holds immense promise for improving the efficiency and precision of document processing and transcription systems, reducing the need for manual intervention, and advancing character recognition in a variety of handwritten documents across different languages and contexts. This research represents a comprehensive and innovative approach to addressing the enduring challenges of handwritten character recognition and its applications in the realms of Deep Learning, document analysis, and character recognition systems.

Keywords: Handwritten character recognition, Ensemble Learning, Deep Learning, Convolutional Neural Network (CNN), Pattern Recognition, K-Fold cross-validation

Critical Analysis of Various Techniques

The objective of this literature survey is to review the current state-of-the-art in personality prediction systems and explore the benefits of various techniques and approaches. The main goal of this literature survey is to identify the gaps and limitations in existing research and techniques and propose potential solutions to improve the accuracy and robustness of personality prediction models. Additionally, this survey aims to identify the most effective feature extraction and machine learning algorithms for predicting personality traits, as well as exploring the various applications of personality prediction models in different domains such as recruitment, marketing, and mental health.

Advances in Handwritten Character Recognition

Introduction

Handwritten character recognition is a vital area of research with diverse applications, ranging from document analysis to language-specific character recognition systems. Recent studies have made significant strides in this field, capitalizing on ensemble techniques, convolutional neural networks (CNNs), and various preprocessing methods to enhance recognition accuracy and efficiency. This literature survey offers an in-depth overview of the latest developments in handwritten character recognition, encompassing a wide array of techniques, addressing challenges, and exploring their implications across multiple languages.

Ensemble Techniques for Improved Accuracy

Ensemble methods have emerged as a promising approach to substantially enhance the accuracy of handwritten character recognition systems. A notable example is the paper "Handwritten Digit Recognition Using Ensemble Learning" by Kuppa Venkata Padmanabha Nandan et al. (2020). This work leverages ensemble learning in conjunction with CNN models and achieves significant accuracy improvements. The ensemble approach proves particularly effective when dealing with various data splits, such as random and class-wise divisions, underscoring its adaptability and potential to boost recognition accuracy.

Preprocessing: A Cornerstone of Character Recognition

Effective preprocessing techniques play a pivotal role in refining the quality and consistency of input data. "Support Vector Machine Based Handwritten Hindi Character Recognition and Summarization" by Sunil Dhankhar et al. (2021) is a prime illustration of the importance of preprocessing. This study employs a range of methods, including morphological operations, edge detection, and histograms of oriented gradients (HOG), to enhance the recognition of handwritten Hindi characters. These preprocessing steps significantly contribute to achieving remarkable precision in character recognition, emphasizing their critical role in the overall process.

CNNs: The Backbone of Recognition Systems

Convolutional neural networks (CNNs) continue to be a cornerstone in recent research endeavors to attain state-of-the-art results in handwritten character recognition. "Convolutional neural network-based ensemble methods to recognize Bangla handwritten characters" by Mir Moynuddin Ahmed Shibly et al. (2021) exemplifies the powerful impact of CNNs when trained on a Bangla handwritten character dataset. This research explores various CNN architectures and shallow machine learning algorithms, demonstrating the pivotal role of CNNs in character recognition. The high accuracy achieved underscores the proficiency of CNNs in this domain.

Language-Specific Character Recognition

Recent research ventures extend their reach to encompass a wide range of languages, showcasing the extensive applicability of handwritten character recognition. "A Novel Weighted SVM Classifier Based on SCA for Handwritten Marathi Character Recognition" by Surendra P. Ramteke et al. (2019) exemplifies the recognition of Marathi characters. The study introduces a unique approach with a high degree of accuracy. By addressing the challenge of language-specific character recognition, this research paves the way for similar studies focused on under-resourced languages. It underscores the importance of recognizing linguistic diversity in character recognition systems.

Image Preprocessing and Recognition Accuracy

Handling variations in handwritten characters is a substantial challenge in character recognition. "Research on Influence of Image Preprocessing on Handwritten Number Recognition Accuracy" by Tieming Chen et al. (2019) underscores the significance of preprocessing techniques. The study emphasizes the importance of character segmentation, tilt correction, offset correction, and size normalization in elevating recognition accuracy. These techniques yield positive results across various recognition algorithms, including support vector machines (SVMs), hidden Markov models (HMMs), CNNs, and extreme learning machines (ELMs).

Transfer Learning and Robust Recognition

Transfer learning emerges as a potent strategy in the development of character recognition systems. "Recognition of Handwritten Japanese Characters Using Ensemble of Convolutional Neural Networks" by Angel I. Solis et al. (2023) introduces a CNN-ensemble model employing transfer learning to recognize handwritten Kanji characters. This research showcases the model's adaptability to diverse datasets, underlining the importance of recognizing complex, multi-script characters. Transfer learning stands out as a formidable tool for constructing robust recognition systems capable of handling intricate and diverse scripts effectively.

Comparative Study

Comparative Summarization

Table 1 summarizes some of the recent research on HWR. The table includes information on the authors, year of publication, journal, techniques used, and summary of the research. The research covers a variety of topics, including ensemble learning, feature extraction, and pre-processing.

Techniques Used	Summary	Reference
CNN with dropouts	Ensemble learning for handwritten digit recognition using a simple CNN model	[1]
Morphology, HOG, SVM, RSVM, Modified Pihu	Two-stage approach for recognizing handwritten Hindi characters and summarizing text	
CNNs, Ensemble methods (Bagging, Boosting, Random Forests, etc.)	CNNs with different architectures, shallow machine learning algorithms, and ensemble techniques to recognize Bangla handwritten characters	4
Weighted One-Against-Rest Support Vector Machines (WOAR-SVM), Sine Cosine Algorithm (SCA), Morphological operations, Modified Pihu, Statistical, Global, Geometrical, and Topological feature extraction	Weighted SVM classifier with SCA and various feature extraction methods for recognizing handwritten Marathi characters	5
Background modeling, Fingertip detection, Virtual character reconstruction, Normalization, DTW-based classifier	Vision-based finger-writing character recognition system	10
Preprocessing (Tilt correction, Offset correction, Size normalization, Thinning)	Preprocessing algorithm for hand-written character recognition	3
Grayscale, normalization, Statistical Support Vector Machine (SVM)	Statistical SVM-based framework for handwritten character recognition	8
CNN, MLP, Various preprocessing and post-processing techniques	Supervised classifier approach for recognizing handwritten Gujarati characters	6
tilt correction, offset correction, size normalization, and thinning. *SVMs, NNs,HMMs,RFs*	Impact of image preprocessing techniques on recognition accuracy for handwritten numbers	7
CNNs, Transfer learning	CNN-ensemble model for recognizing handwritten Japanese characters (Kanji)	9

Table 1: Summarization of Recent Literature

Comparative accuracy evaluation

Table 2 summarizes the accuracies of various handwritten character recognition (HWR) techniques and algorithms presented in recent research papers.

Paper Title	**Accuracy**
Handwritten Digit Recognition Using Ensemble Learning	99% (Random split), 95% (Class-wise split)
Support Vector Machine Based Handwritten Hindi Character Recognition and Summarization	96.97%
Convolutional Neural Network-based Ensemble Methods to Recognize Bangla Handwritten Characters	98.68% (Ekush dataset), 98.69% (BanglaLekha-Isolated dataset)
A Novel Weighted SVM Classifier Based on SCA for Handwritten Marathi Character Recognition	95.14%
A Novel Vision-Based Finger-Writing Character Recognition System	95.3% (Uppercase), 98.7% (Lowercase English alphabets)
Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition	99.70%
Gujarati Handwritten Character Recognition from Text Images	97.21%
Recognition of Handwritten Japanese Characters Using Ensemble of Convolutional Neural Networks	Ranging from 95% to 99.35% on various datasets
Research on Influence of Image Preprocessing on Handwritten Number Recognition Accuracy	Significant improvements for various recognition algorithms
A Preprocessing Algorithm for Hand-Written Character Recognition	Up to 5% Imporvements

Table 2: Summarization of Accuracy and Improvements

References

[1] M. P. a. S. V. Kuppa Venkata Padmanabha Nandan, ""Handwritten Digit Recognition Using Ensemble," *IEEE,* July 27,2020.

[2] M. K. G. F. H. M. S. B. D. a. A. M. Sunil Dhankhar, "Support Vector Machine Based Handwritten Hindi Character Recognition and Summarization," *Tech Science Press,* 08-Nov-21.

[3] A. S. A. M. W.H. ABDULLA, "A preprocessing algorithm for hand-written character recognition," *Elsevier Science Publishers,* 02-Oct-17.

[4] T. A. T. "Mir Moynuddin Ahmed Shibly*, ""Convolutional neural network-based," *PeerJ,* 28-Jun-21.

[5] A. A. G. &. D. S. D. Surendra P. Ramteke, ""A Novel Weighted SVM Classifier Based on SCA for," *IETE Journal of Research,* 19 Jun 2019..

[6]   . A. &. M. R. A. &. M. Jyoti Pareek (Ph.D), "Gujarati Handwritten Character Recognition from Text Images," *Procedia Computer Science, Elsevier,* 04-Jun-20.

[7] G. F. H. W. a. Y. L. Tieming Chen(&), "Research on Influence of Image Preprocessing on Handwritten Number Recognition Accuracy," *Springer, Cham,* 13-Apr-19.

[8] P. S. Yasir Babiker Hamdan, "Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition," *Journal of Information Technology and Digital World,* 08-06-2021.

[9] J. Z. J. L. a. A. A. Angel I. Solis, "Recognition of Handwritten Japanese Characters Using Ensemble of Convolutional Neural Network," *ARXIV,* 06-Jun-23.

[10] L.-X. Z. a. J.-C. H. LIANWEN JIN and DUANDUAN YANG, ""A NOVEL VISION-BASED FINGER-WRITING CHARACTER," *World Scientific,* 20-Sep-06.