Tamil Handwritten Character Recognition: A Comprehensive Review of Recent Innovations and Progress

Manoj K, Iyapparaja M
School of Computer Science, Engineering and Information Systems Vellore Institute of Technology, Vellore, Tamil Nadu, India 635810

***Abstract*—Recognizing handwritten characters is a complex task, particularly when dealing with Tamil, a writing system known for its intricate and stylized nature. Several challenges arise in recognizing Tamil handwritten characters, including the complexity of the writing style, similar character shapes, irregular handwriting, slanting characters, varying curves, inconsistent font sizes, and limited datasets. Additionally, the diversity of writing styles and the absence of a standard solution for accurately recognizing all Tamil characters further complicate the process. To address these issues, researchers have explored various techniques, including neural networks, support vector machines, clustering, and groupwise classification. However, Tamil handwritten character recognition remains an evolving field with ample opportunities for exploration and advancement. This review study aims to provide a thorough analysis of the current state of the field, identify key challenges, and highlight areas for improvement. Furthermore, it presents a detailed examination of the proposed techniques and suggests potential directions for future research in this domain.**

***Keywords*—Convolutional Neural Network (CNN); handwritten recognition; Tamil characters; offline recognition; feature extraction techniques; neural network architecture; Support Vector Machine (SVM); groupwise classification**

I. INTRODUCTION

Handwritten recognition is a crucial aspect of computer vision and has various practical applications such as reading handwritten documents, name recognition, postal address recognition and so on [1]. The process of recognizing Tamil handwritten characters has been a topic of interest in the field of computer vision and pattern recognition [2]. Due to the increasing trend of digitalization, there is an escalating need for precise and effective techniques that can identify handwritten characters and convert them into a digital text format [3]. Identifying Tamil characters, especially in handwritten form, is a difficult task as it involves various hurdles, such as slanted characters, similar-looking characters, connected characters, and curved shapes [4]. However, these techniques have their own limitations and may not provide optimal accuracy and efficiency. Multiple methods have been used to tackle the challenges in the recognition of handwritten characters, including groupwise classification, clustering, support vector machine (SVM), and the Component labeling method. However, these methods have limitations in terms of accuracy and efficiency [5].

Recently, Convolutional Neural Network (CNN) has emerged as a promising solution for handwritten recognition as it can effectively classify and recognize characters [6]. Unlike traditional feature extraction techniques, the system being proposed uses the pixel values obtained from the resized characters during the segmentation stage, which results in a less complex system [7]. In recent times, Artificial Neural Network (ANN) has been used for handwritten recognition. This approach has shown promising results in terms of accuracy and speed [8]. The process of ANN for handwritten recognition involves various stages such as segmentation, feature extraction, and classification. The pixel values obtained from the resized characters during the segmentation stage are utilized to train the neural network. The precision of the recognition system is dependent on the quantity of hidden layers and the number of neurons present in each layer [9]. Moreover, the proposed system is designed to recognize a minimal character set and has limitations in terms of different writing styles and font sizes [10]. However, the system can be further improved in the future by exploring the key challenges in this field, as shown in Fig. 1. The proposed system has the potential to find applications in handwritten name recognition, postal address recognition, and document reading. Through this, our aim is to make a contribution to the progress of research in the field of handwritten recognition and bring us a step closer to a standard solution for identifying Tamil characters with a satisfactory degree of accuracy.

This study is structured as follows: In Section I, an introduction is provided. Section II focuses on the motivation behind the study. Section III highlights the summarized literature review. Table I compares and contrasts different studies. Section IV discusses the challenges in handwritten Tamil character recognition. In Section V, the study concludes by summarizing the findings and outcomes of the study.

A. Tamil Handwriting Recognition

Tamil Handwriting recognition is a field of study that aims to develop algorithms and systems capable of recognizing the handwritten Tamil script [11]. With the increasing amount of digital data and the need for efficient data processing [12], the importance of recognizing and processing handwritten data has become increasingly important. The Tamil script is a complex script, with a large number of characters and variations in handwriting styles, making the recognition process challenging [13]. Numerous

algorithms and approaches have been developed over the years to tackle this challenge. These include traditional methods like clustering and groupwise classification, support vector machines (SVM), Component labeling methods, and Artificial Neural Networks (ANN). While these methods have provided some level of accuracy in recognizing Tamil handwriting, they have certain limitations, such as being time-consuming and complex [14]. The architecture for Tamil handwritten character recognition (HCR) typically involves several stages, including preprocessing, segmentation, feature extraction, and recognition. During the preprocessing stage, the input image is transformed into a form that is suitable for processing, such as converting it to grayscale or thresholding it [15]. The segmentation stage is used to separate individual characters from the background. Feature extraction techniques are employed to obtain significant features from the segmented characters that can be used to represent them. The recognition stage is where the extracted features are utilized to classify the characters and identify their identity [16]. In summary, Tamil Handwriting recognition remains a challenging and ongoing field of study, with ongoing research aimed at improving the accuracy and efficiency of recognition algorithms. With advanced techniques such as deep learning and CNNs, there is great potential for further improvements in this field, which has applications in various areas such as document reading, handwritten name recognition, and postal address recognition.

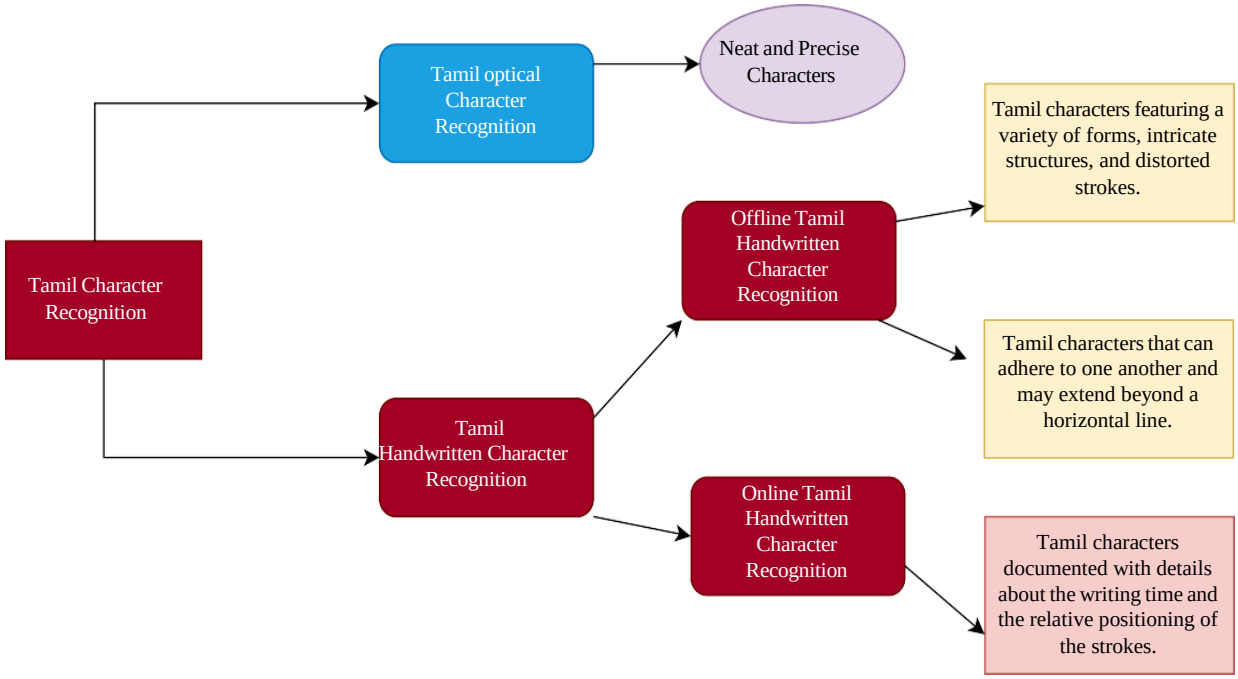

Fig. 1. Classification of Tamil character recognition.

TABLE I COMPARATIVE STUDY

S.No	Methodology	Achievement	Limitations	Ref No
I	Optimized (CNN)	The proposed CNN + SALA model achieves an accuracy of 0.84 at TD=50, outperforming RNN, LA, EHO-NN, and DCNN by 12.7%, 3.4%, 7.2%, and 11.2%, respectively.	The results reveal that although the proposed TCR system performs better than existing models in terms of accuracy, sensitivity, specificity, and precision, it still needs further improvements to achieve optimal recognition of Tamil characters.	[30]
II	MNM (NN+EHO)	Modified traditional neural network that utilizes Elephant Herding Optimization for optimizing weights for the identification of Tamil characters. The MNN method demonstrates an impressive recognition rate of 93%, surpassing the results achieved by other existing methods such as NN, SOM, SVM, FNN, RBF, and Quad tree.	Only Tamil character recognition. Limited comparison to existing classifiers. Evaluation based only on recognition rate.	[19]
III	ResNet (type of ANN with multiple layer skips and nonlinearities, including batch normalization)	Proposed system for Tamil character recognition has 96% accuracy, using a dataset of over 15,000 images of 128x128 dimensions.	Limited to recognition of only 128x128 images of Tamil characters [31]. Similar characters may cause recognition errors. Limited to Tamil character recognition only, cannot recognize words or sentences.	[32]
IV	Support Vector Machines (SVM)	The system has obtained 82.04% recognition accu-racy on handwritten Tamil characters.	Trained with only 35,441 characters from 106 distinct Tamil characters handwritten by 117 individuals. Recognition accuracy varies	[22]

			for different Tamil characters, ranging from 62.84% to 98.9%.	
V	A modified convolutional neural network (M-CNN)	The system achieved a high recognition accuracy of 97.07% for the recognized characters using the proposed method.	The work is limited to isolated characters and there is scope for future work on recognizing Tamil handwritten words and sentences.	[33]
VI	Java Neural Network	Achieved 100% recognition rate with 0.1 second recognition speed and 97% accuracy on the test set.	Only 25 users' handwriting was used for training, with 15 users for testing. Recognition is limited to handwritten Tamil characters from 25 selected users.	[34]
VII	Hidden Markov Models (HMM)	The system's performance for retrieving handwritten documents is 80.75%.	The system has been designed and tested for the Tamil language only. Effectiveness on documents with complex layouts and multiple font styles, as well as handling degraded or damaged images, is limited.	[35]
VIII	SVM + Modified Quadratic Discriminant Function (MQDF)	Evaluated on 36,172 data points, achieving an accuracy of 95.13%.	Limited sample size. Limited feature consideration. Limited script coverage. Limited error analysis. No real-world testing.	[36]
IX	SVM with Feature Selection Algorithms	Achieved 89% accuracy for testing samples.	Optimized for cursive characters, which may result in errors if the writing style varies significantly. Limited feature selection and statistical-based feature extraction algorithms.	[37]
X	SVM, ELM, and MLP Classifiers	The highest accuracy of 95.26% was achieved by the MVTI-SVM combination.	Comparison limited to 4 feature extraction methods. Structural features could enhance recognition accuracy, particularly for ambiguous character pairs.	[38]
XI	ADAM Optimizer	CNN + ADAM achieved 92% accuracy, CNN + Lion Optimization achieved 86%, RMSProp achieved 89%, and AdaGrad achieved 88%.	Processing time is high with the ADAM Optimizer.	[39]
XII	Convolutional Neural Networks	The best neural network architecture for character recognition had two hidden layers of 100 neurons each, with 90.19% accuracy.	Does not cover different writing styles and font sizes. Limited character set. Neural network architecture may not suit other writing styles.	[40]
XIII	Convolutional Neural Networks	A proposed system uses a CNN to recognize 70 Tamil characters with 93% accuracy, outperforming methods like Clustering, SVM, Component labeling, and ANN.	The dataset is limited to 200-250 samples per character from the hp-labs-Tamil-dataset. Only online datasets with specified class labels are supported. Recognizes a limited number of characters.	[41]

II. Motivation

The motivation behind researching handwritten character recognition in Tamil lies in the growing need for fast and accurate recognition of handwritten characters in various domains such as healthcare, transportation, finance, and education. Despite being a complex task, the application of handwritten character recognition in these domains can bring about significant improvements and streamline processes. The challenge arises from the intricate variations in writing styles and individual differences among writers, making traditional recognition methods inadequate. To overcome these difficulties, researchers have been exploring the use of various algorithms for handwritten character recognition. The parallel architecture and fault-tolerance have demonstrated promising results in addressing the intricacies involved in character recognition. The primary objective of this research study is to provide a comprehensive overview of the current state-of-the-art techniques and methodologies used in the domain of Tamil handwritten character recognition. Additionally, this study seeks to identify areas for future research and contribute to the development of advanced and dependable systems for recognizing handwritten characters in Tamil.

III. Summarized Literature Review

Kavitha et al. [17] presents a study on recognizing offline handwritten Tamil characters through the use of Convolutional Neural Networks (CNNs). The study aims to establish a standard for offline Handwritten Tamil Character Recognition (HTCR) by leveraging deep learning techniques. The researchers utilized an isolated handwritten Tamil character dataset that was developed by HP Labs India and trained the model from scratch, achieving good recognition results with a training accuracy of 95.16%. The Tamil Language has a unique feature of having a syllable for every sound pronounced, making the word economy of characters minimal. The Tamil script comprises a total of 307 characters, but the complete Tamil character set can be represented by only 156 distinct characters. This study is significant because it establishes the standard for HTCR for all 156 classes through the application of deep learning methods, unlike previous studies, which used a limited number of classes or used grouping methods.

The field of optical character recognition (OCR) has gained momentum in recent years due to the growing need to convert scanned images into machine-readable formats. The OCR process finds applications in various fields, such as intelligent scanning machines, language translators, and text-to-speech converters. The development of a Tamil character recognition (TCR) system is one of the challenging tasks in OCR, primarily due to the unique problems that the Tamil language poses to the developers. The existing TCR systems face challenges such as variations in the size and style of

handwriting, orientation angle, and font size [18].

To overcome these challenges, Kowsalya et al. [19] proposed an efficient Tamil character recognition system that involves four primary processes: preprocessing, segmentation, feature extraction, and recognition. The preprocessing process includes Gaussian filtering, binarization, and skew detection. The segmentation process involves line and character segmentation. The extracted features are then used for recognition, which is carried out using an optimal artificial neural network. To optimize the weights of the neural network, the researchers used the Elephant Herding Optimization algorithm. The performance of the proposed method was evaluated using sensitivity, accuracy metrics, and specificity. The proposed method is implemented in MATLAB and has shown promising results in the analysis. Previous research has focused on developing efficient TCR systems, and significant progress has indeed been made in the field of Tamil character recognition over the years. However, there are still challenges related to existing character recognition methods, such as the loss of information and change in shape after binarization. The proposed method aims to address these challenges by incorporating effective preprocessing and a modified neural network. The proposed method has the potential to serve as a solution for the development of an accurate Tamil character recognition system. The objective of the paper by Harini et al. [20] is to digitize old handwritten Tamil scripts through the use of Optical Character Recognition (OCR) technology. Tamil is a classical language of India with a rich literary tradition spanning over four thousand years, with much of its literature still undigitized. The OCR system in this study uses four main modules: Scanning, Preprocessing, Recognition, Feature Extraction, and Image Segmentation. The input is a scanned image of handwritten Tamil scripts, which undergoes processes like binarization, noise removal, skew correction, segmentation, and feature extraction to produce a digital output. This module compares the input sample to the trained dataset and generates a digital version of the handwritten input as output.

Raj et al. [21] review pertains to a study on a Tamil handwritten character recognition system. The primary goal of this study is to extract and identify distinct features of different Tamil characters that exhibit variations in their shapes and styles. Three methods of feature extraction were experimented with: shape, location-based instances, and shape ordering. The proposed system consists of four phases: pre-feature extraction, feature extraction, Z-ordering, and classification. The feature extraction phase involves strip tree, Z-ordering, and PM- Quad tree algorithms. The strip tree algorithm deals with shape features, the Z-ordering algorithm is used to address the problem of structure ordering, while the PM-Quad tree is used to facilitate efficient searching and retrieval of characters. The classification phase uses a hierarchical support vector machine to predict characters based on a divide-and-conquer procedure. The researchers collected data for the experiment from the HP-India Handwritten Tamil Character Dataset and individuals of different age groups. They utilized a total of 10,000 characters, with 6,000 characters being used for training and 4,000 characters for testing purposes. The experiment was implemented in MATLAB, and the results show that SVM performed better than other classifiers. The study concludes that the proposed system can recognize a larger number of characters and their varied shapes.

Shanthi et al. [22] aim to recognize offline handwritten Tamil characters using a support vector machine (SVM). The study uses handwritten characters collected from various writers on A4-sized documents, scanned at 300 dpi resolution and stored as grayscale images. Pre-processing is done on the digitized images to improve quality, and pixel densities are computed for 64 different zones to form character features used for SVM training. The system achieves an 82.04% recognition accuracy on the handwritten Tamil character database.

Sornam et al. [23] focused on introducing a new approach to recognize offline handwritten Tamil characters, which is challenging due to their complexity and variability in writing style among different users. Their method involves a combination of principal component analysis (PCA) and a convolutional neural network (CNN) for feature extraction and recognition of Tamil characters, as shown in Fig. 2, resulting in a more accurate recognition system. Deep learning, a state-of-the-art technique in machine learning [24], was used to reduce the burden on the programmer and to extract reliable features for complex problems such as object recognition and handwritten character recognition. The dataset utilized for the research was gathered from diverse writers located in Bangalore and Salem, India. It included both offline and online versions of Tamil characters. The model was trained for 50 epochs and achieved a convergence of approximately 88% in the 27th epoch. The training accuracy increased over the 50 epochs, and the performance measures were promising. The study presented a Tamil character recognition system that showed good performance. The system has the potential to be further extended for recognizing slogans in images with the use of a GPU.

The article by Jose et al. [25] discusses a wavelet transform-based feature extraction technique for recognizing Tamil handwritten characters. The classification process in the algorithm is carried out using a feed-forward back propagation neural network, while feature extraction is done through two-dimensional discrete wavelet transformations. The article highlights the challenges in recognizing Tamil characters, including the large number of characters and complex handwriting styles, and notes that preprocessing, segmentation, feature extraction, and classification are the major steps involved in character recognition. The proposed system has achieved a maximum recognition rate of 89% after preprocessing, wavelet decomposition, and BPNN classification.

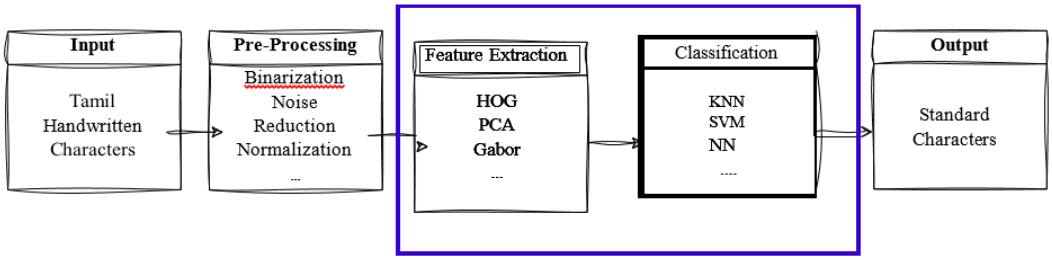

Fig. 2. Traditional offline HTCR process and typical methods for each step.

Wahi et al. [26] concentrate on creating an Optical Character Recognition (OCR) system to recognize handwritten Tamil characters. The study highlights the use of Zernike moments and Legendre polynomials for feature extraction of handwritten Tamil characters and a feed-forward backpropagation neural classifier for character classification. The proposed system consists of three modules: preprocessing, classification and feature extraction. The preprocessing module involves the conversion of color images to binary images through binarization techniques such as thresholding and Otsu's method. The feature extraction module includes the computation of various character properties such as height, width, horizontal lines, vertical lines, slope lines, circles, and arcs. The research concludes that there is a lack of studies on recognizing handwritten characters of Indian scripts, and a complete recognition system is not yet available in India due to the extensive character set of Indian languages, the presence of vowel modifiers and compound characters. The study focuses on the recognition of handwritten South Indian languages, namely Tamil.

Shaffi et al. [27] presents the development of an extensive Tamil Handwritten Character Database (uTHCD) for research in document image analysis. The database consists of 91000 samples in 156 character classes with around 600 samples in each class, collected from 650 volunteers, including school children, teachers, homemakers, faculty and university students. The samples were collected both offline and online and encompass a wide range of writing styles and inherent distortions. The authors anticipate that the database will establish a new standard in research on Tamil handwritten character recognition and provide a starting point for other projects in document image analysis. The study also outlines an experiment conducted on the database using a Convolutional Neural Network (CNN) architecture, which achieved an initial accuracy of 88% on the test data. The existing Tamil Handwritten Character Database, developed by HP Labs India, contains 156 distinct character classes with a varying number of samples in each class. Ad-hoc databases have also been created by various researchers, but the sample size is limited and varies greatly between different works. The uTHCD was created through a sample collection process starting with volunteers writing characters in a 10x8 grid on an A4 sized paper or an electronic form using a digital writing pad. Preprocessing was done to remove unnecessary information and to enhance the quality of the samples. The authors aim to make the database publicly available in raw image and HDF compressed format for easy implementation.

Rajasekar et al. [28] focus on the development of a computer-based handwritten character recognition (HCR) system for Tamil characters. The historical background and development of optical character recognition (OCR) and HCR systems in general are discussed. The article provides an overview of prior handwritten character recognition (HCR) methods and techniques, including those that employ moment-based feature extraction and coding schemes based on neighborhood relations. Data reduction is performed using the fuzzy c-means clustering algorithm, and recognition is carried out using Artificial Neural Networks (ANN). The proposed system in the paper involves three modules: preprocessing, segmentation, and recognition. The preprocessing module includes image filtering and binarization, as well as skew correction. The segmentation module decomposes the image into individual characters. The recognition phase involves the use of correlation and neural network methods. The objective of this study is to create a MATLAB program using the image processing toolbox to develop an OCR system for Tamil alphanumeric and text.

This literature review by Suganthe et al. [29] focuses on the development of a handwritten text recognition (HTR) model for the Tamil language. The study highlights the importance of feature extraction in HTR models and explains how Convolutional Neural Networks (CNNs) can be used for automatic feature extraction. The literature review outlines the steps involved in digitizing handwritten text recognition, including the development of a CNN model, hyper-parameter tuning, character segmentation, and recognition of text by feeding individual characters to the tuned model. The Tamil language is considered as a classical language and has a rich history, with a unique writing system that includes 247 characters, with 156 distinct characters capable of depicting the entire Tamil character set. However, handwritten text recognition for Tamil characters is a challenging task due to various distortions, variations in character length, wide range of handwriting styles, fragmented strokes, and noise. The preprocessing steps involved in the HTR model for Tamil characters include grayscale conversion, binary conversion, foreground character extraction, morphological transformation, line segmentation, and character segmentation. The hp-labs- tamil-iso-character dataset was used in this study and the architecture of the CNN model is discussed in detail, including the use of convolutional and pooling layers. The study concludes with a discussion of the experimental results and future scope of the research.

The proposed literature review by Kishan et al. [42] focuses on the field of Handwritten Character Recognition

(HCR) using Convolutional Neural Networks (CNN). Handwritten character recognition (HCR) is an active and challenging area of research in image processing and pattern recognition, with applications in reading bank cheques and recognizing characters from form applications. The proposed work extends the EMNIST dataset by adding characters from the Tamil language, and trains the CNN using this extended dataset. The feature extraction process includes converting the input image to a grayscale image, normalizing the pixel values, and transforming the input data into a set of features through dimensionality reduction. The Min-Max-Scaler is used to normalize the data, and the CNN is used as a classifier. The system includes preprocessing, feature extraction, Min-Max-Scaler, image normalization, and classification in its architecture. The preprocessing step involves converting the input image to a grayscale image and resizing it to match the resolution of the EMNIST dataset. The process of feature extraction involves transforming the input data into a smaller set of features that can effectively represent the input data. The Min-Max-Scaler normalizes the data to a range between 0 and 1, and the trained CNN is used to recognize the characters and provide the desired output.

Thilagavathi et al. [43] focus on the proposed system for the conversion of handwritten South Indian language characters, specifically Tamil, into digital characters. The learning methodology of the system involves the use of Artificial Neural Networks (ANNs) with the Stochastic Gradient Learning Algorithm and Back-Propagation. Previous implementations of similar systems have shown accuracy rates over 90%. The system architecture includes pre-processing modules for image to pixel conversion, a network trainer, and a character predictor. The dataset used for the system design and implementation is from a worldwide competition on handwritten character recognition. The study also highlights the importance of character recognition in various industries, including banking and subjective answer evaluation, as well as the challenges posed by the Tamil language with its wide use of symbols.

Livingston et al. [44] have developed a handwritten Tamil character recognition system that uses a combination of Hidden Markov Model (HMM) and Fuzzy Logic. The system aims to convert handwritten Tamil characters into Unicode and is a challenging task due to the great variation among handprints. Among the different branches of character recognition, Tamil character recognition is more complicated due to multiple patterns to represent a single character and disconnected multi-stroke characters. The authors propose a framework for handwritten Tamil character recognition using a combination of Hidden Markov Model and Fuzzy Logic. The framework consists of three stages: preprocessing, feature extraction using Hidden Markov Model, and classification using Fuzzy Logic. The Fuzzy technique is used to differentiate between Tamil scripts. The system was tested on 750 samples of Tamil characters and achieved a success rate of 89 to 93%. The authors note that there has been significant research in the field of handwriting recognition, including the use of neural networks and hidden Markov models. While the most popular class of nonlinear vector operators in handwritten recognition is based on order statistics, the authors propose a different approach.

Gayathri et al. [45] conducted research on the recognition of cursive characters in Tamil palm leaf manuscripts using the Convolutional Neural Network (CNN) technique. The authors used a text line slicing segmentation scheme to segment the palm leaf characters, and morphological operations were used for preprocessing to eliminate noise in the input image. The feature processing included text line spacing, spacing without obstacles, and spacing with obstacles. The proposed network achieved 94% accuracy, while the existing ResNet achieved 88% accuracy. The proposed methodology is divided into four steps: i) collection of cursive Tamil palm leaf manuscripts, ii) preprocessing, iii) feature extraction, and iv) classification using CNN. The results and discussion section of the study shows that the proposed CNN model outperforms the current models in terms of accuracy.

The literature review by Suriya et al. [46] focuses on the application of Computational Linguistics and Convolutional Neural Network in the recognition of Tamil characters. The study presents an Intellectual Character Recognition System which uses CNN to recognize the Tamil character dataset and achieved a training accuracy of 99.16%. The challenges in handwritten character recognition are also discussed, including the variations in writing style, difficulties in recognizing similar characters, etc. The study proposed a Character Recognition system using a Convolutional Neural Network and used a dataset developed by HP Labs India, which consists of 156 unique Tamil characters written by native Tamil writers. The study highlights the use of Data Augmentation to improve the variety of data available for training models and overcome the issue of limited data size. The literature review concludes that the proposed system has significant potential in the field of handwritten character recognition, as it is capable of recognizing characters in challenging conditions where traditional character recognition systems fail. The comparative study highlights various methodologies for Tamil handwritten character recognition, with CNN- based approaches achieving the highest accuracy, such as 97.07% using Modified CNNs and 96% with ResNet. Hybrid methods combining CNNs with optimization techniques, like Elephant Herding Optimization, also demonstrated strong performance, achieving up to 93% accuracy. Traditional methods like SVM and HMM showed moderate success, with accuracies ranging from 80.75% to 95.13%, but are limited by smaller datasets and simpler feature extraction techniques. Despite advancements, most methods are restricted to isolated character recognition and rely on small, constrained datasets, limiting generalizability as shown in Table I. The bar chart in Fig. 3 underscores the superior accuracy of CNN-based models, indicating the need for future work on recognizing entire words or sentences and handling diverse datasets.

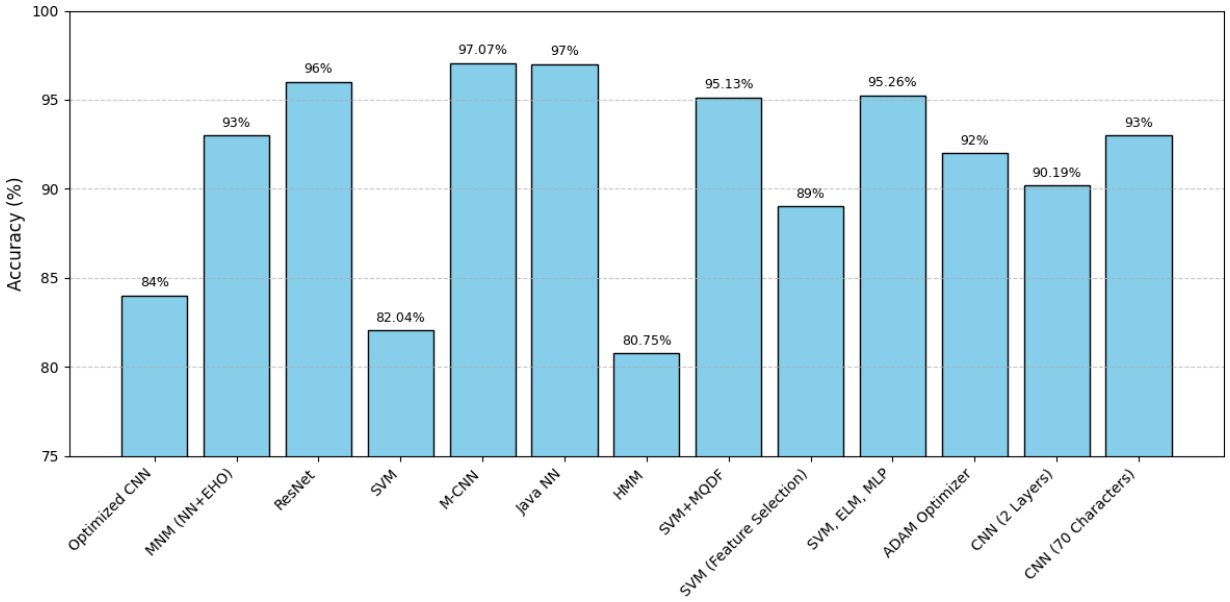

Fig. 3. Comparison of accuracy across methods for Tamil HCR.

IV. Challenges in HCR Tamil

The Tamil writing style can be complex, including variations such as normal and abnormal writing, slanted characters, similar-shaped characters, curves, joined characters, and more, which makes it difficult to accurately recognize the characters [47]. There are many Tamil characters that have similar shapes, which can lead to confusion during recognition [48]. Handwritten characters can be written in an abnormal style, making it challenging to accurately recognize them with standard recognition algorithms [49]. The slanting of characters can also affect recognition, as it can change the character's shape and make it look different from the standard form [50]. Tamil characters can contain curves, which can also pose a challenge for accurate recognition [51]. Different people write in different font sizes, and this variation can pose a challenge for accurate recognition [52]. Feature extraction is often used in recognition processes, but this method can be complex and time-consuming [53]. There is a lack of large datasets for Tamil handwritten character recognition, which limits the development of effective recognition algorithms [54]. People write in different styles, and it can be difficult to accurately recognize characters. In Fig. 4, the Tamil characters have been written in a different style than the one used to train the recognition system [55]. Despite numerous studies and efforts in the field of Tamil handwritten character recognition, there is still a lack of a universally accepted and efficient solution for accurately identifying all Tamil characters [56]. To overcome these challenges, researchers have proposed various approaches including the use of neural networks, support vector machines, clustering, and groupwise classification. The field is still evolving and there is room for further improvement.

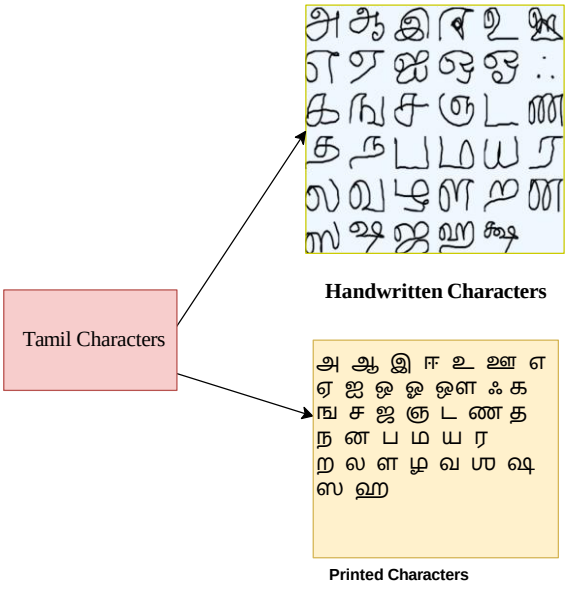

Fig. 4. Comparison of optical vs. handwritten.

V. Conclusion

In conclusion, the research on handwritten character recognition (HCR) in Tamil has made significant strides in recent years. A multitude of algorithms and techniques have been proposed and implemented to tackle the challenges of recognizing handwritten characters in the Tamil script. From the reviewed studies, it can be deduced that Convolutional Neural Networks (CNNs) have emerged as a prominent solution for HCR in Tamil due to their high accuracy and ability to effectively capture the inherent features of handwritten characters. However, there is still room for improvement and further research in this field to enhance the performance and efficiency of HCR systems for Tamil script. This includes the optimization of feature extraction techniques, the development of more robust neural network architectures, and the integration of other advanced machine

learning algorithms. In light of these developments, it is clear that the field of HCR for Tamil script holds immense potential for innovation and advancement, and will continue to play a crucial role in various domains and applications

REFERENCES

[1] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari, and Gulshan Kumar. A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27:1071–1092, 2020.

[2] B. Radhika. An empirical analysis of Tamil optical character recognition. *Journal of Innovative Image Processing*, 4(2):115– 126, 2022.

[3] S. D. Pande, P. P. Jadhav, R. Joshi, A. D. Sawant, V. Muddebihalkar, S. Rathod, and S. Das. Digitization of handwritten de- vanagari text using cnn transfer learning-a better customer service support. *Neuroscience Informatics*, 2(3):100016–100016, 2022.

[4] K. Subramani and M. Subramaniam. Creation of original Tamil character dataset through segregation of ancient palm leaf manuscripts in medicine. *Expert Systems*, 38(1):12538–12538, 2021.

[5] C Vinotheni and S Lakshmana Pandian. End-to-end deep- learning-based Tamil handwritten document recognition and classification model, 2023.

[6] A. Morera, A. Sa´nchez, J. F. Ve´lez, and A. B. Moreno. Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, pages 1–14, 2018.

[7] D. S. Prashanth, R. V. K. Mehta, K. Ramana, and V. Bhaskar. Handwritten devanagari character recognition using modified lenet and alexnet convolution neural networks. *Wireless Personal Communications*, 122:349–378, 2022.

[8] J. K. Basu, D. Bhattacharyya, and T. H. Kim. Use of artificial neural network in pattern recognition. *International Journal of Software Engineering and Its Applications*, 4(2), 2010.

[9] M. M. Misgar, F. Mushtaq, S. S. Khurana, and M. Kumar. Recognition of offline handwritten urdu characters using rnn and lstm models. *Multimedia Tools and Applications*, pages 1–24, 2022.

[10] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. volume 39, pages 2298–2304. IEEE, 2016.

[11] E. Vellingiriraj, M. Balamurugan, and P. Balasubramanie. Text analysis and information retrieval of historical tamil ancient documents using machine translation in image zoning. *Inter- national Journal of Language Literature Linguistics*, 2(4):164– 168, 2016.

[12] M. Iyapparaja and P. Sivakumar. Detecting diabetic retinopa- thy exudates in digital image processing hybrid methodology. *Research Journal of Pharmacy and Technology*, 12(1):57–61, 2019.

[13] S. Akashkumar, A. N. Dyaram, and M. Anand. Identification of tamil characters using deep learning. In *Machine Learning and Autonomous Systems: Proceedings of ICMLAS 2021*, pages 223–237. Springer Nature, 2022.

[14] S. Vijayalakshmi, K. R. Kavitha, B. Saravanan, R. Ajaybaskar, and M. Makesh. Handwritten character recognition for tamil language using convolutional recurrent neural network. In *Inventive Systems and Control: Proceedings of ICISC 2022*, pages 369–384. Springer Nature, 2022.

[15] M. A. Mohamad, H. Hassan, D. Nasien, and H. Haron. A review on feature extraction and feature selection for handwritten character recognition. *International Journal of Advanced Computer Science and Applications*, 6(2), 2015.

[16] N. Rani, T. Shobha, M. Vasudev, N. Chandrajith, and Manohar. 2d morphable feature space for handwritten character recognition. *Procedia Computer Science*, 167:2276–2285, 2020.

[17] B Rose Kavitha and C b Srimathi. Benchmarking on offline handwritten tamil character recognition using convolutional neural networks, 2022.

[18] A. K. Sampath and N. Gomathi. Decision tree and deep learning based probabilistic model for character recognition. *Journal of Central South University*, 24:2862–2876, 2017.

[19] S. Kowsalya and P. S. Periasamy. Recognition of tamil handwritten character using modified neural network with aid of elephant herding optimization. *Multimedia Tools and Applications*, 78:25043–25061, 2019.

[20] ML Harini, A Pandian, A Hithayathullah, and AK Sadiq. Handwritten tamil character recognition.

[21] M. A. R. Raj and S. Abirami. Structural representation- based off-line tamil handwritten character recognition. *Soft Computing*, 24:1447–1472, 2020.

[22] N. Shanthi and K. Duraiswamy. A novel svm-based handwritten tamil character recognition system. *Pattern Analysis and Applications*, 13:173–180, 2010.

[23] M. Sornam and C. V. Priya. Deep convolutional neural network for handwritten tamil character recognition using principal component analysis. In *Smart and Innovative Trends in Next Generation Computing Technologies: Third International Conference*, pages 778–787. Springer, 2017.

[24] C. Gopalakrishnan and M. Iyapparaja. Multilevel-thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning. *International Journal of System Assurance Engineering and Management*, pages 1–8, 2021.

[25] T. M. Jose and A. Wahi. Recognition of tamil handwritten characters using daubechies wavelet transforms and feed-forward backpropagation network. *International Journal of Computer Applications*, 64(8):975–8887, 2013.

[26] A. Wahi, S. Sundaramurthy, and P. Poovizhi. Handwritten tamil character recognition using zernike moments and legendre polynomial. In *Artificial Intelligence and Evolutionary Algo- rithms in Engineering Systems: Proceedings of ICAEES 2014*, volume 2, pages 595–603. Springer, 2015.

[27] N. Shaffi and F. Hajamohideen. uthcd: a new benchmarking for tamil handwritten ocr. *IEEE Access*, 9:101469–101493, 2021.

[28] M. Rajasekar, A. Bennet, and M. Celine. Performance and analysis of handwritten tamil character recognition using artifi- cial neural networks. *International Journal of Recent Scientific Research*, 7(1):8611–8615, 2016.

[29] R. C. Suganthe, K. Pavithra, N. Shanthi, and R. S. Latha. A cnn model based approach for offline handwritten tamil text recognition system. *NVEO-Natural Volatiles Essential Oils Journal*, pages 164–175, 2021.

[30] R. B. Lincy and R. Gayathri. Optimally configured convolu- tional neural network for tamil handwritten character recogni- tion by improved lion optimization model. *Multimedia Tools and Applications*, 80:5917–5943, 2021.

[31] N. Sankarram B. S. Charulatha R. Jayakanthan, A. H. Kumar and A. Ramesh. Handwritten tamil character recognition using resnet. *International Journal of Research in Engineering*, 3(3):133–137, 2020.

[32] C. Gopalakrishnan and M. Iyapparaja. Detection of polycystic ovary syndrome from ultrasound images using sift descriptors. *Bonfring International Journal of Software Engineering and Soft Computing*, 9(2):26–30, 2019.

[33] S. L. Pandian C. Vinotheni and G. Lakshmi. Modified con- volutional neural network of tamil character recognition. In *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2020*, pages 469–480. Springer, 2020.

[34] C. Sureshkumar and T. Ravichandran. Handwritten tamil character recognition and conversion using neural network. *International Journal on Computer Science and Engineering*, 2(07):2261–2267, 2010.

[35] S. Palanivel A. N. Sigappi and V. Ramalingam. Handwritten document retrieval system for tamil language. *International Journal of Computer Application*, pages 31–31, 2011.

[36] P. Ka and P. Elangob. Accuracy improvement of off-line handwritten tamil character recognition. *Malaya Journal of Matematik (MJM)*, (2):504–512, 2015.

[37] M. A. R. Raj and S. Abirami. Offline tamil handwritten character recognition using statistical features. *Advances in Natural and Applied Sciences*, (9):367–375, 2015.

[38] S. S. Mari and G. Raju. Modified view-based approaches for

handwritten tamil character recognition. *ICTACT Journal on Image Video Processing*, 6(1), 2015.

[39] P. Jagathratchagan R. Iyswarya, S. Deepak and J. Kailash. Handwritten tamil character recognition using convolution neu- ral network by adam optimizer. *International Journal of Advanced Research in Science and Communication Technology (IJARSCT)*, 6(1), 2021.

[40] R. Meena M. Deepa, R. Deepa and R. Nandhini. Tamil hand- written text recognition using convolutional neural networks. *International Journal of Engineering Science and Computing*, 9(3):20 986–20 988, 2019.

[41] A. Sivamurugan S. Ramkumar, M. Saivignesh and K. Shan- mugam. Digitalization of tamil handwritten characters recog- nition using convolutional neural networks (cnn). *International Journal of Advance Research and Innovative Ideas in Educa- tion*, 6:262–268, 2020.

[42] S Anandh Kishan, J Clinton David, P Sharon Femi, and Sri Venkateswara. Handwritten character recognition using cnn. *International Journal of Research and Analytical Reviews*, 5(3), 2018.

[43] G. Lavanya G. Thilagavathi and N. K. Karthikeyan. Tamil handwritten character recognition using artificial neural net- work. *International Journal of Scientific Technology Research*, 8, 2019.

[44] L. G. X. Livingston and L. M. M. Livingston. Handwritten recognition of tamil scripts using hmm with fuzzy logic. *International Journal of Scientific Research*, 2017.

[45] Y. Teekaraman R. Kuppusamy S. G. Devi, S. Vairavasundaram and A. Radhakrishnan. A deep learning approach for recog- nizing the cursive tamil characters in palm leaf manuscripts. *Computational Intelligence And Neuroscience*, 2022.

[46] S. Dhivya S. Suriya and M. Balaji. Intelligent character recognition system using convolutional neural network. *EAI Endorsed Transactions on Cloud Systems*, 6, 2019.

[47] KAKND Dharmapala, WPMV Wijesooriya, CP Chandrasekara, UKAU Rathnapriya, and L Ranathunga. Sinhala handwriting recognition mechanism using zone based feature extraction, 2015.

[48] V. P. Singh H. Singh, R. K. Sharma and M. Kumar. Recognition of online handwritten gurmukhi characters using recurrent neural network classifier. *Soft Computing*, 25:6329–6338, 2021.

[49] A. U. Rahman and Z. Halim. A graph-based solution for writer identification from handwritten text. *Knowledge and Information Systems*, 64(6):1501–1523, 2022.

[50] P. Badelia S. Ghosh A. Chaudhuri, K. Mandaviya. Handwritten text recognition methods and techniques. In *Applications of Pattern Recognition*, pages 9–41. 2017.

[51] T. Deselaers H. A. Rowley A. Daryin M. Calvo Gervais V. Car- bune, P. Gonnet and P. Fast multi-language lstm-based online handwriting recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 23(2):89–102, 2020.

[52] X. He S. Long and C. Yao. Scene text detection and recogni- tion: The deep learning era. *International Journal of Computer Vision*, 129:161–184, 2021.

[53] L. Fernandes S. Santos M. Abreu P. Bota M. Barandas, D. Fol- gado and H. Gamboa. Tsfel: Time series feature extraction library. *SoftwareX*, page 100456, 2020.

[54] K Shanmugam and B Vanathi. Humming bird optimization based deep belief neural network for tamil handwritten charac- ter recognition, 2024.

[55] R. D. Zarro and M. A. Anwer. Recognition-based online kurdish character recognition using hidden markov model and harmony search. *Engineering Science and Technology, an International Journal*, 20(2):783–794, 2017.

[56] A. Bharath and S. Madhvanath. Handwritten document recog- nition: Challenges and advancements. In *Handbook of Pattern Recognition and Computer Vision*, pages 209–234. 2010.