



## A Survey on character Recognition based on Deep Learning techniques

V.Kumaresan<sup>1</sup>, S.Jeya Mary<sup>2</sup>

1. Assistant Professor, Department of ECE, St.Michael College of Engineering & Technology, Kalayarkoil

2. Assistant Professor, Department of ECE, AR College of Engineering and Technology, Kadayam

**Abstract :** This study has been undertaken to investigate the Tamil language characters by utilizing various Deep learning methods. RCNN is a sort of Artificial Neural Network (ANN) which is regularly comprises of twofold or triple layers skirts that has nonlinearities and clump standardization. Tamil is one among most conventional dialects particularly found on the southern districts of India. Perceiving a non-digitalized character in Tamil is an extremely challenging cycle in view of its huge and compound person set. In this paper RCNN Architecture which clarifies the RCNN layers and the trouble in acquiring the digitalized character from the RCNN Layer utilizing profound learning. This includes the informational collection preparing and pre-handling of all the Tamil characters. There are absolutely 256 characters in Tamil language in which the vast majority of the letters are practically comparable and just a slight change can be seen distinctly toward the end for a large portion of the characters, so perceiving a specific person is very troublesome and require exceptionally prepared datasets for recognizing every single characters.

**IndexTerms - Component,formatting,style,styling,insert.**

### I. INTRODUCTION

Tamil is one of the standard Indian dialects which is dominantly used in Southern India. The forte of Tamil language is each strong articulated elements a syllable in Tamil. The economy of characters to address a word is negligible in Tamil language. The littlest unit of Tamil content is syllable. These syllabic units of Tamil content has 12 vowels, 18 consonants and an extraordinary person Ayudha Ezhuthu(ஐ). There are 247 characters, among that 206 compound characters are shaped by vowels and consonants. There are around 5 acquired consonants from Sanskrit, when these sanskrit consonants joined with tamil vowels would yield another 60 compound characters so on make a loaded up with 307 characters.

### II. RELATED WORKS

M. A. Pragathi, K. Priyadarshini, S. Saveetha, A. Shavar Banu, K. O. Mohammed Aarif "Manually written Tamil Character Recognition Using Deep Learning" 2019 [2]-Character acknowledgment is produced for different examples of transcribed or optical characters to be perceived carefully. There are numerous Tamil literary works in undigitized structure. Utilizing profound learning the undigitized Tamil written works can be changed over into meaningful arrangement. Many explores were continued person acknowledgment utilizing profound learning for dialects like Arabic, Devanagari, Telugu, and so on... Due to the bigger class set and disarray in similitudes between manually written characters Tamil person acknowledgment is a test. In this paper, we propose a person acknowledgment framework for manually written Tamil characters utilizing profound learning. Here, VGG 16 methodologies is done. The proposed work gives proficiency of 94.52% on our datasets.

U. Bhattacharya, S. K. Ghosh and S. K. Parui "A Two Stage Recognition Scheme for Handwritten Tamil Characters" 2007 [1] India is a multilingual multiscrypt country with more than 18 dialects and 10 distinctive major contents. Insufficient exploration pursue acknowledgment of transcribed characters of these Indian contents has been finished. Tamil, an official just as famous content of the southern piece of India, Singapore, Malaysia, and Sri Lanka has an enormous person set which incorporates many compound characters. A couple of pursues penmanship acknowledgment of this huge person set has been accounted for in the writing. As of late, HP Labs India fostered a data set of written by hand Tamil characters. In the current paper, we portray a disconnected acknowledgment approach dependent on this information base. The proposed strategy comprises of two phases. In the first stage, we apply an unaided bunching strategy to make fewer gatherings of written by hand Tamil person classes. In the subsequent stage, we consider a regulated classification method in each of these more modest gatherings for final acknowledgment. The provisions considered in the two phases are unique. The proposed two-stage acknowledgment conspire gave satisfactory classification exactnesses on both the preparation and test sets of the current data set.

Manigandan T, Vidhya V, Dhanalakshmi V, Nirmala B "Tamil Character Recognition from Ancient Epigraphical Inscription utilizing OCR and NLP" 2017 [3]-Recognition of antiquated Tamil characters is one of the difficult undertaking for Epigraphers as the language has developed with various characters set. On the off chance that the engravings are on stone dividers, it adds significantly greater intricacy in recognizing characters. This proposed work mostly centers around acknowledgment of different Tamil characters somewhere in the range of ninth and twelfth hundreds of years utilizing OCR and NLP procedures. In this work, the engraving pictures gathered from Tamil Nadu, Archeological Department are pre-handled and divided. During the division cycle the shading pictures were changed over to dark picture and to twofold picture dependent on edge esteem. From fragmented, picture highlights like number of lines, bends, circles and spots have been extricated utilizing Scale Invariant Feature Transform (SIFT) calculations for each letter to distinguish the specific person. Characters will be characterized and built dependent on Vectors separated, utilizing Support Vector Machine (SVM) classifier and the examples of the person will be coordinated with known characters and anticipated utilizing Trigram method. Each distinguished person will be appointed with its comparing Unicode worth and it will be refreshed in the picture corpus for additional person ID, and to make the framework in recognizing the characters all the more successfully. Accordingly the proposed framework can take care of the serious issues in perusing the engraving pictures.

N. Prameela, P. Anjusha, R. Karthik "Disconnected Telugu Handwritten Characters Recognition utilizing optical person acknowledgment" 2017 [4]-The Aim of the proposed paper is to perceive disconnected Hand composed Telugu characters utilizing Optical person acknowledgment, OCR is one of the most famous and testing subject of example acknowledgment This paper proposes an OCR framework for Telugu archives which includes three phases, in particular pre-handling, highlight extraction, and order. In the preprocessing stage, we have utilized middle separating on the information characters and applied standardization and skeletonization strategy over characters for extraction of limit edge pixel focuses. In the component extraction stage, at first each character is isolated into  $3 \times 3$  matrices and the relating centroid for all the nine zones are assessed. With this we can recognize the characters of various styles. From that point, we have drawn the flat and vertical symmetric projection holy messenger to the closest pixel of the person which is named as Binary External Symmetry Axis Constellation for unconstrained written by hand character. From which we have determined the even and vertical Euclidean distance for the equivalent closest pixel from centroid of each zone. Then, at that point, we have determined the mean Euclidean distance just as the mean precise upsides of the zones. This is considered as the key component upsides of our proposed framework. Finally, both help vector machine (SVM) and Quadratic segregate Classifier (QDA) has been independently utilized as the classifier.

Weerasinghe. "Fostering a business grade Tamil OCR for perceiving textual style and size autonomous text" 2015[5]-Optical Character Recognition (OCR) of Indic scripts, for example, Tamil and Sinhala has fallen behind those for dialects dependent on the Latin content. A few endeavors to fabricate business grade OCR for these dialects have fizzled in the past attributable to them not summing up well. This paper portrays a bunch of preparing systems for Tamil utilizing the Tesseract motor that have empowered us to foster a vigorous Tamil OCR framework. We depict exhaustively our preparation system, which brings about an exhibition improvement of 12.5 % over the default Tamil module transported with Tesseract on a bunch of old Tamil reports, which were essential for a credible venture to digitize significant Tamil compositions of Sri Lanka.

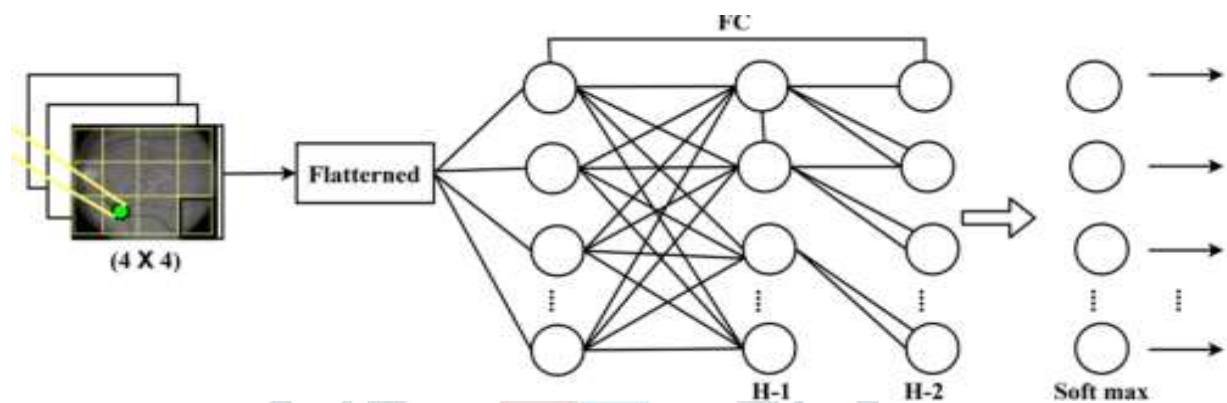
There are a wide range of approaches accessible for character acknowledgment framework. Probably the most normally utilized strategies are support vector machines(SVM), CNN utilizing VGG-16, CNN utilizing LeNet, CNN utilizing GoogleNet, etc. This paper unmistakably clarifies the approach of utilizing Residual Neural Network(ResNet). In view of execute the various approaches the proficiency and precision of framework is examined. On conquering every one of those detriments in the above techniques, this functioning model can ready to create an exactness of 96% which is a lot of higher when contrasted with execution of different frameworks. The accompanying table shows the exactness boundaries for various proposed frameworks.

TABLE I:-PERFORMANCE ANALYSIS

METHODOLGY	ACCURACY	SENSITIVITY	SPECITIVITY
Resnet	99.2%	98.3%	98.0%
3D CNNs	97.52%	95.31%	99.73%
marker-controlled watershed technique	93.23%	93.26%	93.2%
multi-view convolutional networks (ConvNets)	93.7%	95.5%	94.28%
nonlinear algorithm	98.8%	97.7%	96.2%

## 2.1 R-Convolutional Neural Network (CNN)

CNN may be a multistage feedforward artificial neural network with trainable supervised learning. The convolution operation is multidimensional. during a convolutional network, the primary parameter is typically called an input, and therefore the second parameter is named a kernel function, and therefore the output is named a feature map. Sparse representations (also referred to as sparse weights), parameter sharing, and isomorphic representations are three important architectural ideas of CNN. Traditional neural networks use matrix operation to affect connection relationships. An output unit is related to each input unit, which inevitably requires tons of storage. However, the character of the sparse representation of the convolutional network and therefore the neurons are only connected to many neurons adjacent to the previous stage, and therefore the local convolution operation is performed, which reduces the storage requirements and improves the computational efficiency. CNN's parameter sharing abandons the non uniqueness of weights in traditional networks. The weights within the CNN stage are constant, which is best than others in storage requirements. Traditional automatic encoders are fully connected. Vector output and source image aren't necessarily aligned in space, while U-Net uses local connection structure. Vector output and source image are aligned in space, therefore the visual effect of fusion image is best . U-Net may be a full-convolution network [16], which consists of contraction path and expansion path. In-depth learning training needs an outsized number of samples, while U-Net is improved supported full convolution neural network, and may train alittle number of samples using data enhancement. This advantage just caters to the shortcoming of alittle sample size of medical image data.



## III.CONCLUSION

In this venture we perceived transcribed tamil characters ,by giving information picture and afterward it is changed over to a computerized text design. This advanced text can be put away in a report record ,and furthermore a preparation and test precision of 93% is acquired by this CNN model which is superior to Clustering and groupwise arrangement, support vector machine,Component marking technique and ANN.

## REFERENCES

- [1] U. Bhattacharya, S. K. Ghosh and S. K. Parui "A Two Stage Recognition Scheme for Handwritten Tamil Characters" IEEE Transactions on Ninth International Conference on Document Analysis and Recognition, 2007.
- [2] M. A. Pragathi, K. Priyadarshini, S. Saveetha, A. Shavar Banu, K. O. Mohammed Aarif "Handwritten Tamil Character Recognition Using Deep Learning" IEEE Transactions on International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019.
- [3] Manigandan T, Vidhya V, Dhanalakshmi V, Nirmala B "Tamil Character Recognition from Ancient Epigraphical Inscription using OCR and NLP" IEEE Transactions on International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).
- [4] N. Prameela, P. Anjusha, and R. Karthik "Off-line Telugu Handwritten Characters Recognition using optical character recognition" IEEE Transactions on International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017.

- [5] C. S. Sundaresan, S. S. Keerthi, "A Study of Representations for Pen based Handwriting Recognition of Tamil Characters," IEEE Transactions on 1999.
- [6] Chamila Liyanage, Thilini Nadungodage, and Ruvan Weerasinghe. "Developing a commercial grade Tamil OCR for recognizing font and size independent text," IEEE Transactions on International Conference on Advances in ICT for Emerging Regions (ICTer), 2015.
- [7] K. B. M. R. Batuwita, G. E. M. D. C. Bandara "Fuzzy Recognition of Offline Handwritten Numeric Characters" 2006.
- [8] R. Ramanathan, S. Ponmathavan, N. Valliappan L. Thaneshwaran, Arun.
- S. Nair, K. P. Soman "Optical Character Recognition for English and Tamil Using Support Vector Machines" IEEE Transactions on International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009.
- [9] B. S. Charulatha, Paul Rodrigues, T. Chitralekha, Arun Rajaraman "Mining Ambiguities using Pixel-based Content Extraction", Advances in Intelligent Systems and Computing, pp. 537 – 544, 2016.
- [10] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang. Phoneme recognition using time delay neural networks. IEEE Transactions on Acoustics, Speech and Signal Processing, March 1989.
- [11] Shaohan Xu, Qi Wu, and Siyuan Zhang "Application of Neural Network in Handwriting Recognition" IEEE Transactions on International Conference of Stanford University.

