A Comparative Analysis for Optical Character Recognition for Text Extraction from Images Using Artificial Neural Network Fuzzy Inference System

Srikanth Bhyrapuneni*, Anandan Rajendran

Department of Computer Science and Engineering, VELS Institute of Science, Technology and Advanced Studies, Chennai 600117, India

Corresponding Author Email: srikanthbt3010@gmail.com

ABSTRACT

Artificial neural networks (ANN) has the capability to analyze raw data from processing input-output relationships. This function considers them important in areas of industry with such information is unusual. Researchers have tried to extract the information embedded within ANNs as set of rules used with inference systems to resolve the black-box function of ANNs. When ANN applied within a fuzzy inference system, the extracted rules yield high classification accuracy. In this paper a Multi-Layer Neural Feed-Forward Network using Artificial Neural Network Fuzzy Inference System (MLNFFN-ANNFIS) is proposed for accurate character recognition from images. The technique targets areas of business that have less complicated issues about which there is no simpler approach is desired to a complex one. This paper proposed an Optical Character Recognition model for Text Extraction from Images using Artificial Neural Network Fuzzy Inference System for accurate text detection from images. The technique proposed is more effective and simple than most of the techniques previously proposed. The proposed model is compared with various traditional models and the results indicate that the proposed model accuracy is more and performance is also improved.

1. INTRODUCTION

ANN systems are computational systems of low-level, parallel-processing that are shown to be universal approximates for processing raw data to approximate any input-output relationship [1-4]. The process of text extraction begins by scanning and correctly recognizing the characters, and after identification, the data is translated to computer-readable format. Computer or digital machine can easily use it but a human need time [5]. There are a variety of ways of doing this. Artificial intelligence systems have started recognizing photos and faces better. Artificial neural network and various data mining techniques are used for character recognition [6].

Optical Character Recognition (OCR) has proved exceedingly difficult due to texture and contrast of nature pictures. On the picture, since some patterns are visually indistinguishable from the true text, it may be incomprehensible [7]. These images have various colors, sizes, forms, orientation and types of fonts, often curvy. Those photos often suffer a significant amount of errors including exposure, noise, blur, underexposed, changing exposure and so on. Recognition of text from the images is a challenging task and the proposed model introduced an Artificial neural network based Deep learning model for accurate text detection from images [8].

The Fuzzy inference system is designed to have eight inputs that correspond to eight template matrix pixels and one output [9]. The input image is transformed into the fuzzy domain [0 1] at the point of fuzzification [10]. The edge image is transformed back to the image pixel domain at the interval [1 255] during the defuzzification process. In order to compare them with the edge detection method, the PSNR error metric is predicted and the proposed fuzzy edge detection method is found to be accurate. To create an expert inference system, these rules could then be used [11]. In fact, some research indicates that sometimes the collection of extracted rules will outperform the generalization of the trained artificial neural network from which the rules have been extracted [12].

Research has spread in several directions to derive rules from a qualified artificial neural network. For such algorithms, the main classification scheme is in the way they extract rules. Each hidden and output node is individually analyzed in the decomposition approach, and a rule is extracted from it for accurate text detection from images. The output of each neuron generated in a feed-forward neural network is measured as:

$$O_j = \left(\left(\sum_i w_{ij} \times I_i \right) + \theta_j \right) \tag{1}$$

where, $Ii = \frac{1}{1+e^{-ax}}$.

Here, I is the activation status of the neuron i, wij is the weight of the relation between the neuron i and the neuron j, θ is the activation degree of the neuron j that regulates the steepness of the sigmoid function. The most significant aspect of the decomposition method is that there are roughly 0 or 1 activations of all neurons in the ANN. In the neurons of the hidden layer, binary inputs cause this to happen.

1.1 Deep neural networks

The Deep Neural Network (DNN) consists of processors called neurons that are interconnected. Data is processed by each neuron and passed on to the next neurons. Similar to the human brain, it was designed to function [13]. The proposed model uses DNN for accurate and fast text recognition from Images [14]. The frequency of the association between the two units of information is directly proportional to the interaction between them. The neural network depicts the data that the machine understands [14]. DNN has been used in the area of machine learning to solve different tasks that are difficult to solve, such as speech recognition, handwriting recognition, and question-answering. DNN can be used to store previously processed information and it can be used for knowledge inference later on. The general procedure for Optical Character Recognition is depicted in Figure 1.
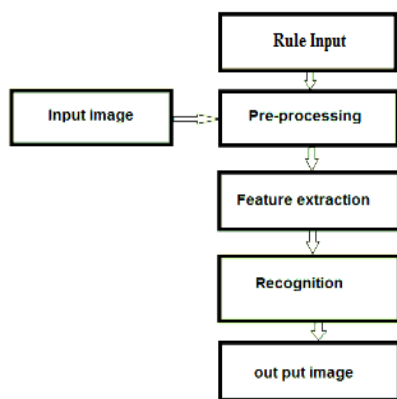
Figure 1. Optical character recognition procedure

NLP is an area in computational linguistics, cognitive science and artificial intelligence. In human-computer interaction, it plays an important role [15]. A machine only knows 1's and 0's, and by using natural language, humans communicate with each other [16]. The language used by human NLP is needed for a computer to comprehend and process [17]. DNN includes different layers that are secret. Such layers can help define the link between different units of information when generating a response. In a simple ANN, two neurons may not seem to be directly linked to each other but are associated indirectly by DNN. If we create a DNN for the same thing, we can find some indirect relations.

2. LITERATURE SURVEY

The Key phrases from text algorithm was introduced several separate rule-extracting algorithms have been proposed in the past two decades [1]. where, for each secret and output neuron, it looks for a single relation with a sufficiently large weight to exceed the neuron bias. The algorithm searches for two relation subsets that outweigh the bias, followed by a three-part subset, and so on. When all the neurons have been searched, rules extracted from the neurons of the hidden layer are paired with rules extracted from the neurons of the output layer to establish rules for the input-output relationship to classify text and to delete rules subsumed by the more general rules.

Drobac et al. [3] suggested an updated technique using rule-based fuzzy logic. A fuzzy-based method was built to detect edges without setting the threshold value. The edge of the grey scale image is calculated. Along with the current Sobel and prewitt edge detector, this work suggested soft computing approach-based image edge detection that improves edge detection where the histogram is implemented with fuzzy logic. Eger et al. [5] suggested edge detection in matlab using fuzzy logic. This work is focused on sixteen rules that distinguish between the target pixels. The Fuzzy inference method is combined with conventional edge operators. The noise is extracted at various processing speeds. The first and second derivatives are added to the resulting FIS image.

Evershed and Fitch [7] proposed a model in which heuristic rules were applied and a fuzzy-based approach was used to simultaneously perform filtering and edge extraction. A fuzzy edge detection filter is used that goes through two phases of the process to eliminate noise from grayscale images. SWT is used to classify linked components and to detect text positions through unsupervised clustering. A method that recognizes texts regardless of their scale, direction, font and size is suggested in this model. The approach is based on corner points in a text. A Laplacian approach is used to detect video text that handles text from any direction.

Graves et al. [9] suggested an automated system to identify vehicle registered numbers plates. The input is translated to a binary format. After the initial step of conversion, noise reduction is critical. Sources of noises from car are scratches on registration plate, emissions from exhaust and so on. After the preprocessing stages, four programs were built to create an encoded form from segment-emoji characters. In the first algorithm, the vector crossing is present. It is one of the famous strategies used for the OCR. The main goal of the vector-crossing technique is to count the number of vectors that move through each pixel of the image. In order to get a perfect result two vectors are considered. The vectors will all lie in the center of the image.

Hämäläinen and Hengchen [11] implemented an OCR method to recognize handwritten letters and numerals. In the preprocessing stage, binarization, noise reduction, scaling and normalization and thinning of the images are performed. Both structural and global features are extracted in feature extraction point. This is an attempt to classify the structural features of the picture by splitting it into four quadrants. From each axis 5 features are estimated and all of a sum total of 20 features are estimated in an image. The 20 constituent features plus one feature, which is called the width feature, is regarded as an image. Various global features are extracted with Wavelet transform. Wavelets are mathematical functions that are used to subdivide data into different frequencies and analysis each piece of data with a resolution to fit its size.

A point feature matching approach has been developed to recognize text from the image. The image is processed and correlated with the objects in the database. The item template is a combination of single letter as well as a combination of strokes. In this process, the feature points in the target images will be identified. Extracted features now locate the object in the image.

Kauppinen [13] proposed an OCR system with fast training neural network. In each recognition level, however, was allocated a training period of short length. Training data is now further split into different subgroups. This ranking system is based on factors like symmetry, Euler number features and so on. Neural networks may have applications during the training phase. This would strengthen the machine so it will learn stuff faster. This approach does not involve any preprocessing.

Each single character is passed through a neural network and selected from two possible outcomes for the next step. Many OCR techniques render the problems when doing the character classification in the preprocessing stage. This would help remove the repetitive terms. This similarity search is based on comparison of pixels.

The feature extraction process and the image to matrix mapping process are performed in the model developed by Kettunen and Koistinen [15]. The images are transformed into a 2D matrix. The next move would be to change the computer system. The proposed method used the Multi-Layer Perceptron Learning Algorithm to optimize the decision making. This approach uses a pyramid structure to facilitate learning. This approach can be used not only for classification purposes but also for learning purposes. Applying the learning process algorithm within the multilayer neural network architecture, the synaptic weights and threshold are modified in a way that the processing task can be carried out efficiently. These learning rules are essential for learning objectives. In each iteration, the weight is incremented to some integer value. In order to recognize the object, it figures out features of the object.

Wick et al. [18] proposed a model in which the template matching and back propagation algorithm is introduced. Template matching is a common method used in optical character recognition algorithms. This is perhaps the most well-used feature extraction technique. The system's simplicity makes it more popular among users. Correlation is another expression for template matching. In this technique each single cell is used to perform feature extraction. A data table is compiled and stored in the database. The character nearest to the source term is now chosen as the best match. Back propagation algorithm that uses reverse mechanism to find the error and retransfer errors are used in this model.

3. PROPOSED METHODOLOGY

Optical Character Recognition (OCR) process translates the text into digits. It is a program that analyzes and retrieves information from an image [19]. The image can be a photographic image, text document, or printed picture. Computer OCR converts the images into a readable form of text. The proposed Multi-Layer Neural Feed-Forward Network using Artificial Neural Network Fuzzy Inference System (MLNFFN-ANNFIS) recognizes only the particular characters for which the classification system has been trained using the process.

Optical Character Recognition is the procedure of translating text into a machine-readable text by computer [20]. The textual input may be scanned, picture of paper or from subtitles superimposed on an image. There are three phases of the proposed model namely, learning [21], layer consideration and extraction. When new data is fed to it the machine learns the rules and feature level extraction techniques to identify the relevant pixels in the given image. New neurons (words) or connections between neurons are added to the existing network when learning so that the hidden multi layers [22] are updated and the accuracy in the text recognition will be improved [23].

To extract the pixels from the image, the input data and output data were fed into the multi-layer ANFIS model. The 'fuzzification' layer is set and the parameters for the selected membership are adapted. Initially preprocessing is applied to remove the noisy values from the image. After that features are extracted by considering multiple hidden layers. There is another adaptation layer before the final layer that acts as a 'defuzzification' layer of guidelines, where the parameters of the model are tuned to derive the best fit between input and output.

The proposed model considers the image for text extraction and the process of recognition of text using the relevant pixels are indicted in Figure 2.
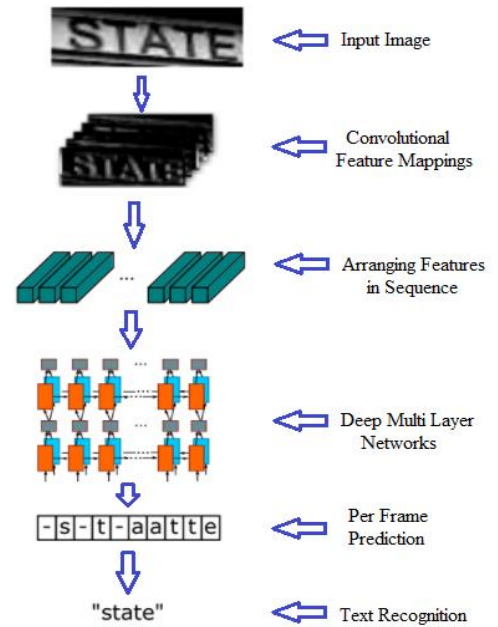

Figure 2. Text extraction process

In order to enhance its inference capabilities, deep cases with multi layers are allocated to the neurons in the artificial neural network [24]. The concise statistical details found in the image are reflected by a photo histogram. Now consider that $I(P, Q)$ is the function representing each pixel's degree of intensity in the image (P, Q), where $P=1, 2, ..., M$ and $Q=1, 2, ..., N$.

Each individual pixel is required for measuring the Grey-level histogram. For each occurring pixel intensity level of 0,1, the probability density 'PD' is determined, N-1 dividing them by the total number of pixels with $G(i)$.

It is possible to portray that as:

$$PD(i) = \frac{h(i)}{MN} + Th + In(i) \ \ for \ \ i = 0,1,......, N-1 \qquad (2)$$

where, $h(i)$ is the histogram value calculated, Th is the threshold limit and In is the intensity level of the image considered. The histogram levels are calculated as:

$$h(i) = \sum_{x=1}^{A} \sum_{Y=1}^{B} \delta(I(P,Q),i) + \theta \qquad (3)$$

here, $h(i)$ is the histogram function for the whole image and for each intensity level δ for image I and θ represents the angle of the image Here we have taken $\delta(i, j)$ of a pixel and its neighbor as the delta function that can be given as:

$$\delta(i, j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \qquad (4)$$

After Calculating the image intensity and histogram levels, the image considered will undergo segmentation. The process of segmentation is performed as:

$$lMG_\sigma(I[l_M, l_N])$$
$$= \sqrt{\sum_{l=l_i}^{M} PD(i) + (T - l_i(I[l_i, l_j]))^2 \times P_l^{I[l_i, l_j]} + \theta} \qquad (5)$$

The proposed work considers multi layers for pixel extraction for accurate text extraction from the image. A single layer is generated for every set of pixels. A single layer is calculated for every i^{th} pixel set. An adaptive pixel set for a feature for an instant layer is calculated as:

$$O_i = \mu Pi(i) + \delta(i, j) \quad for \quad i = 1.2...M$$
$$O_i = \mu Qi(j) - 2(x) + \delta(i, j) \quad for \quad j = 3.4...N \qquad (6)$$

here, μ is the layer priority considered for arranging the layers in a sequential process for accurate and unique text recognition. The single layer generated is a fixed one with a total amount that calculates the total output as the number of all pixels that forms as a set. The proposed model generates multi layers for every set of pixels for a feature representation. A new layer is generated as:

$$O_i = \sum \overline{w}i.hi = \frac{\sum \overline{w}i.h(i) + \theta}{\sum \overline{w}i} + \delta(j, k)fi$$
$$for \quad k = 1, 2, ...M - N \qquad (7)$$

If i and j are the two pixels extracted from a image, the membership grades of $I(i, j)$ pixel are calculated for the image I. The difference between the $M(P)$ and $M(Q)$ pixels can be calculated based on entropy of fuzzy:

$$Dn(P, Q) = \sum_{i=o}^{M}\sum_{j=0}^{N} \begin{bmatrix} \frac{\mu_A^2(M_{ij}) + \mu_N^2(Q_{ij})}{\mu_A(N_{ij}) + \mu_M(P_{ij})} \\ + \frac{(1 - \mu_P(M_{ij}))^2 + (1 - \mu Q(N_{ij}))^2}{2 - \mu_N(P_{ij}) - \mu_M(Q_{ij})} \\ + \theta(P, Q) + Th \end{bmatrix} \qquad (8)$$

Similarly, the similarity level between the calculated entropy levels is calculated to establish the training model of the features for text recognition. The training model is established as:

$$D_1(B, A) = \sum_{i=o}^{n}\sum_{j=0}^{n} \begin{bmatrix} \sqrt{\frac{\mu_A^2(a_{ij}) + \mu_B^2(b_{ij})}{2}} \\ + \sqrt{\frac{1}{M-1}\sum_{i=1}^{N}|P_{i-\mu}|^2} \\ + \sqrt{\frac{(1 - \mu_A(a_{ij}))^2 + (1 - \mu_B(b_{ij}))^2}{2}} + \end{bmatrix} \qquad (9)$$

So, the total divergence between the pixels a_{ij} and b_{ij} of the images A and B due to $m_1(A)$ and $m_1(B)$ is calculated as:

$$Div[m_1(A, B)] = \sum_{i=o}^{n}\sum_{j=0}^{n} \begin{bmatrix} \frac{\mu_A^2(a_{ij}) + \mu_B^2(b_{ij})}{\mu_A(a_{ij}) + \mu_B(b_{ij})} \\ + \frac{(1 - \mu_A(a_{ij}))^2 + (1 - \mu_B(b_{ij}))^2}{2 - \mu_A(a_{ij}) - \mu_B(b_{ij})} \end{bmatrix} +$$
$$\sum_{i=o}^{M}\sum_{j=0}^{N} \left[\sqrt{\frac{(1 - \mu_A(P_{ij}))^2 + (1 - \mu_B(Q_{ij}))^2}{2}} \right] \qquad (10)$$

As DMM model is used that considers the hidden layers for accurate text extraction. The input to each hidden layer is provided:

$$H(I(i, j)) = I_{i,j} + \sum F_i * W \qquad (11)$$

where, F is the bias value on the hidden layer and W is the weights between input and hidden layer. Therefore output from each hidden layer is calculated by:

$$O(I(i, j))_i = \frac{\sum_{j=1}^{N}(A_{ij}^m + B_{ij}^n)x_j}{\sum_{i=1}^{m}(A_{ij}^m + B_{ij}^n)} \qquad (12)$$

The pixels that frame the text extracted from the image are represented as:

$$T(X, Y) = \sum_{(i,j)\varepsilon Fs: i \varepsilon w, j \varepsilon T} O(i) + wi + \frac{(i * j)(i, j)}{(H_{i,j}(P, Q))} \qquad (13)$$

The precision and recall are calculated as:

$$Precision = \frac{\sum B_e \in A^M (A_e, T)}{|P + Q|} \qquad (14)$$

$$Recall = \frac{\sum A_t \in B^m (A_t, E)}{P + Q} \qquad (15)$$

4. RESULTS AND DISCUSSION

The proposed model is implemented in python and executed in ANACONDA Spyder. The proposed model considers the data set from URL https://www.kaggle.com/andrewmvd/car-plate-detection and https://www.kaggle.com/konradb/text-recognition-total-text-dataset. The proposed model considers 45376 images for text recognition from images using the proposed Multi-Layer Neural Feed-Forward Network using Artificial Neural Network Fuzzy Inference System (MLNFFN-ANNFIS) model. The new algorithm proposed in this paper addresses the problem of efficiently extracting a limited pixels of text data that yield high accuracy [25]. Two parameters were varied for the experiments: data used to train the networks and the shape of the functions of the membership of the fuzzy inference systems in which the extracted pixels were applied [26]. The data for network training differed between all data points in the data set and only the data points in the data set that did not create a nonlinear distinction between the text class and the noise class [27].

The proposed MLNFFN-ANNFIS model considers a text

image and vehicle number images and considers the hidden layers and perform pixel extraction from them to perform accurate optical character recognition of the input provided. The Figure 3 represents the images considered for the extraction process.

The input image after preprocessing undergoes noise elimination and the text relevant features are only considered. The Figure 4 represents the number identification process.
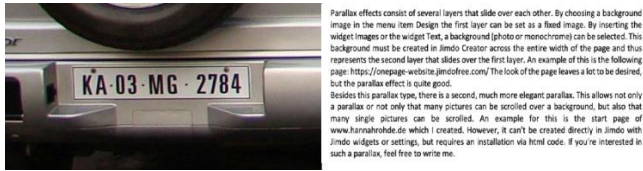

Figure 3. Input images for text extraction

Figure 4. Pixels of number considered

The pixels which form the text accurately are only extracted and they are converted as text. The proposed model uses DNN model for considering hidden layers for improving the accuracy in text extraction from images. The Figure 5 represents the pixel values extracted from the image to form text.

```
[[ 9.9999994e-01   9.9999988e-01]
 [ 5.9604645e-08   0.0000000e+00]
 [ 9.9999994e-01  -9.9999988e-01]
 [ 1.9999998e+00   0.0000000e+00]]
[[ 9.9999994e-01   9.9999988e-01]
 [ 5.9604645e-08   0.0000000e+00]
 [ 9.9999994e-01  -9.9999988e-01]
 [ 1.9999998e+00   0.0000000e+00]]
```

Figure 5. Extracted pixel text

The noisy data is removed from the image and only the boundaries that contains the text are extracted as the multiple layers are responsible in cropping the noisy values and extracts only useful pixel data. The Figure 6 represents the image that is considered from the input image that only contains the text from the image.

Figure 6. Text boundary considered from input

The proposed model considers multi layers and uses Artificial Neural Network Fuzzy Inference System for text extraction from the input provided. The multi layers are used for improving the accuracy of the system. The Figure 7 indicates the extracted text from the image.

Figure 7. Extracted text from input

The Figure 8 represents the process of extracting relevant text from the images. The proposed model effectively extracts only the relevant text and eliminates the noisy data.

Figure 8. Text extraction from image

The proposed model uses multi-layer network ANFIS model for text detection from images. The proposed model text extraction accuracy is compared with the traditional models and the results show that the proposed model is better and the time for text extraction is less. Figure 9 represents the text extraction time levels of the proposed and the traditional methods.
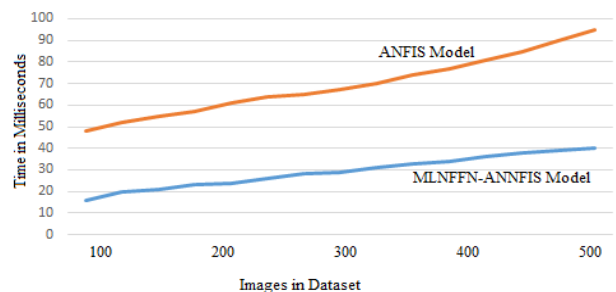

Figure 9. Time for relevant text extraction

In the proposed MLNFFN-ANNFIS model, when a input image is provided, the borders of the image holding the text will only be considered and the remaining part of the image is not considered for the text extraction process. The proposed model efficiently identifies the borders and edges for performing text extraction. Figure 10 describes the accuracy levels of the proposed and traditional models for edge and border detections.

The text detection rate of the proposed and traditional Fuzzy Interference System (FIS) is indicated in Table 1. The number of samples considered and the text recognition rate of the proposed and traditional models are clearly included.

The text recognition accuracy of the proposed and the existing models are illustrated in Figure 11. The accuracy levels indicate that the proposed model accuracy is more when compared to traditional methods.
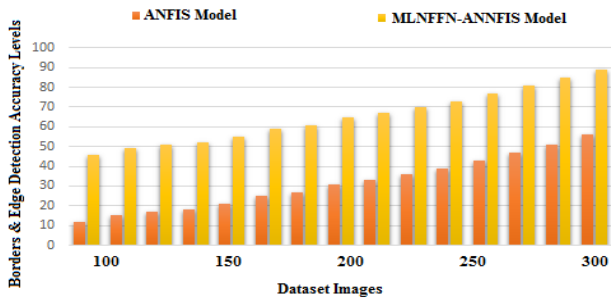
Figure 10. Borders and edge detection accuracy level

Table 1. Test recognition rate with sample range

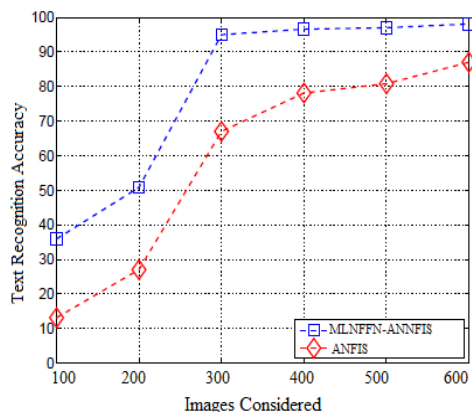| Dataset	Fis		MLNFFN-ANNFIS	
	Number of Recognition Samples	Rate (%)	Number of Recognition Samples	Rate (%)
Training	25,000	77%	25,000	96.5%
Validation	50,000	86%	50,000	95.6%
Test	75,000	74%	75,000	95%

Figure 11. Text recognition accuracy levels

The recall and precision levels of the proposed model and existing models is indicated in Figure 12. The performance of the proposed model is more than the existing method.
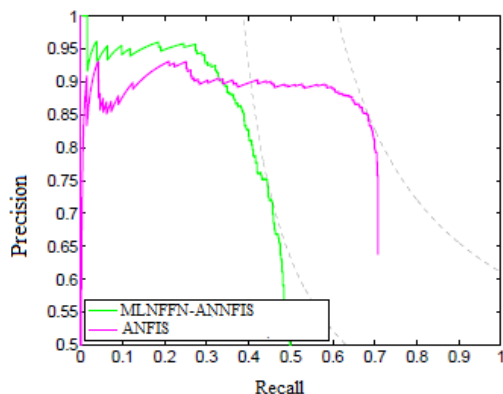

Figure 12. Recall vs precision levels

The proposed model accuracy, sensitivity and specificity levels are indicated in the Table 2. The proposed model is compared to the traditional models and the results show that the proposed model performance is better than the traditional methods.

Table 2. Comparison parameters

Algorithms	Sensitivity	Specificity	Accuracy
MLNFFN-ANNFIS	96.87	94.46	96.35
DWT+KNN	95.1	91.23	94.1
ANN	91.5	92.66	91.3
Texture Combined with ANN	91.6	89.54	87.5
FCM	92%	89.76	89
K-Mean	83%	85.48	78.6

5. CONCLUSIONS

Deep Neural Networks have the ability to learn from simple data processing input-output relationships, a characteristic that often hinders their applicability. Optical Character Recognition is still a challenging task for accurate recognition of text from images. The noise can be removed and that is a positive feature. Sometimes it may not be easy. An important approach to recognizing optical character is discussed in this paper. In this proposed work, for extracting pixels from an image a multi-layer neural feed-forward network for accurate text recognition from images is utilized. The proposed algorithm is capable of extracting general rules that define with high accuracy with the input-output relationships. In addition, the time complexity of O(n) is exhibited by the algorithm, where n is the number of input images. This is much lower than the exponential complexity demonstrated by most traditional algorithms that extract text from images. The proposed solution shows more than 96 percent accuracy. In the future, we would like to find the best soft computing technique for comparing the proposed deep neural network system to its existing systems. In this respect, future work requires changing the algorithm so that it is applicable to multi-layer neural networks. This can be accomplished by increasing the slope of the hidden layer neurons' sigmoid function such that their outputs match the function of binary activation.

REFERENCES

[1] Dong, R., Smith, D.A. (2018). Multi-input attention for unsupervised OCR correction. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2363-2372. http://dx.doi.org/10.18653/v1/P18-1220

[2] Drobac, S., Kauppinen, P., Lindén, K. (2017). OCR and post-correction of historical Finnish texts. In Proceedings of the 21st Nordic Conference on Computational Linguistics, pp. 70-76. https://aclanthology.org/W17-0209.pdf.

[3] Drobac, S., Kauppinen, P., Lindén, K. (2019). Improving OCR of historical newspapers and journals published in Finland. In Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, pp. 97-102. https://doi.org/10.1145/3322905.3322914

[4] Drobac, S., Lindén, K. (2019). Optical font family recognition using a neural network. In Proceedings of the Research Data and Humanities (Rdhum) 2019 Conference Data, Methods and Tools. Studia Humaniora Ouluensia.

[5] Eger, S., vor der Brück, T., Mehler, A. (2016). A comparison of four character-level string-to-string

translation models for (OCR) spelling error correction. The Prague Bulletin of Mathematical Linguistics, 105(1): 77-99. https://doi.org/10.1515/pralin-2016-0001

[6] Englmeier, T., Fink, F., Schulz, K.U. (2019). AI-PoCoTo: Combining automated and interactive OCR postcorrection. In Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, pp. 19-24. https://doi.org/10.1145/3322905.3322908

[7] Evershed, J., Fitch, K. (2014). Correcting noisy OCR: Context beats confusion. In Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, pp. 45-51. https://doi.org/10.1145/2595188.2595200

[8] Généreux, M., Stemle, E.W., Lyding, V., Nicolas, L. (2014). Correcting OCR errors for German in Fraktur font. Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014: 9-11 December 2014, Pisa, 186-190. https://www.torrossa.com/en/resources/an/3044157.

[9] Graves, A., Fernández, S., Gomez, F., Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd International Conference on Machine Learning, pp. 369-376. https://doi.org/10.1145/1143844.1143891

[10] Guha, R., Das, N., Kundu, M., Nasipuri, M., Santosh, K.C. (2020). Devnet: An efficient CNN architecture for handwritten Devanagari character recognition. International Journal of Pattern Recognition and Artificial Intelligence, 34(12): 2052009. https://doi.org/10.1142/S0218001420520096

[11] Hämäläinen, M., Hengchen, S. (2019). From the paft to the fiiture: A fully automatic NMT and word embeddings method for OCR post-correction. In: Recent Advances in Natural Language Processing, pp. 432-437. https://arxiv.org/abs/1910.05535.

[12] Jauhiainen, T.S., Linden, B.K.J., Jauhiainen, H.A. (2016). HeLI, a word-based backoff method for language identification. In Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects VarDial3, Osaka, Japan, December 12 2016. http://hdl.handle.net/10138/174332.

[13] Kauppinen, P. (2016). OCR post-processing by parallel replace rules implemented as weighted finite-state transducers. Doctoral dissertation, University of Helsinki. https://hdl.handle.net/10138/162866.

[14] Kettunen, K.T., Kervinen, J., Koistinen, J.M.O. (2018). Creating and using ground truth OCR sample data for Finnish historical newspapers and journals. In Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference. http://hdl.handle.net/10138/312778.

[15] Kettunen, K., Koistinen, M. (2019). Open source Tesseract in re-OCR of Finnish Fraktur from 19th and early 20th century newspapers and journals-collected notes on quality improvement. In Digital Humanities in the Nordic Countries Proceedings of the Digital Humanities in the Nordic Countries 4th Conference, pp. 270-282. http://hdl.handle.net/10138/311404.

[16] Kissos, I., Dershowitz, N. (2016). OCR error correction using character correction and feature-based word classification. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS), Santorini, Greece, pp. 198-203. https://doi.org/10.1109/DAS.2016.44

[17] Koistinen, M., Kettunen, K., Kervinen, J. (2017). How to improve optical character recognition of historical Finnish newspapers using open source Tesseract OCR engine. In: Proceedings of the LTC, pp. 279-283.

[18] Wick, C., Reul, C., Puppe, F. (2018). Calamari—a high-performance tensorflow-based deep learning package for optical character recognition. arXiv preprint arXiv:1807.02004. https://arxiv.org/abs/1807.02004.

[19] Wick, C., Reul, C., Puppe, F. (2018). Comparison of OCR accuracy on early printed books using the open source engines calamari and OCRopus. J. Lang. Technol. Comput. Linguistics, 33(1): 79-96.

[20] Koistinen, M., Kettunen, K., Pääkkönen, T. (2017). Improving optical character recognition of Finnish historical newspapers with a combination of Fraktur& Antiqua models and image preprocessing. In: Proceedings of the 21st Nordic Conference on Computational Linguistics, pp. 277-283. http://hdl.handle.net/10138/310105.

[21] Reul, C., Springmann, U., Wick, C., Puppe, F. (2018). State of the art optical character recognition of 19th century Fraktur scripts using open source engines. arXiv preprint arXiv:1810.03436. https://arxiv.org/abs/1810.03436.

[22] Farhat, A., Al-Zawqari, A., Al-Qahtani, A., Hommos, O., Bensaali, F., Amira, A., Zhai, X. (2016). OCR based feature extraction and template matching algorithms for Qatari number plate. In 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), Sharjah, United Arab Emirates, pp. 1-5. https://doi.org/10.1109/ICCSII.2016.7462419

[23] Pasha, S., Padma, M.C. (2015). Handwritten Kannada character recognition using wavelet transform and structural features. In 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, pp. 346-351. https://doi.org/10.1109/ERECT.2015.7499039

[24] Bedruz, R.A., Sybingco, E., Quiros, A.R., Uy, A.C., Vicerra, R.R., Dadios, E. (2016). Fuzzy logic based vehicular plate character recognition system using image segmentation and scale-invariant feature transform. In 2016 IEEE Region 10 Conference (TENCON), Singapore, pp. 676-681. https://doi.org/10.1109/TENCON.2016.7848088

[25] Khan, W.Q., Khan, R.Q. (2015). Urdu optical character recognition technique using point feature matching; a generic approach. In 2015 International Conference on Information and Communication Technologies (ICICT), Karachi, Pakistan, pp. 1-7. https://doi.org/10.1109/ICICT.2015.7469576

[26] Lin, H.Y., Hsu, C.Y. (2016). Optical character recognition with fast training neural network. In 2016 IEEE International Conference on Industrial Technology (ICIT), Taipei, Taiwan, pp. 1458-1461. https://doi.org/10.1109/ICIT.2016.7474973

[27] Alam, T. (2015). An approach to empirical optical character recognition paradigm using multi-layer perceptorn neural network. In 2015 18th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 132-137. https://doi.org/10.1109/ICCITechn.2015.7488056