



## RESEARCH ARTICLE

# Intelligent Handwritten Identification Using Novel Hybrid Convolutional Neural Networks – Long-short-term Memory Architecture

Fatimatelbatoul M. Husari, Mohammad A. Assaad

Department of Informatics and Software Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq

**ABSTRACT**

Handwritten character identification finds broad applications in document analysis, digital forensics, and human-computer interaction. Conventional methods encounter challenges in accurately deciphering a range of handwriting styles and variations. Consequently, investigating intelligent systems for handwriting identification becomes crucial to enhance accuracy and efficiency. This paper introduces an innovative hybrid deep learning architecture, seamlessly integrating the strengths of convolutional neural networks (CNN) and long-short-term memory (LSTM) within a single framework. The combination of these structures enables the model to effectively capture both spatial features and temporal dependencies inherent in handwritten strokes, resulting in improved recognition performance. The proposed 2DCNN-LSTM algorithm has been tested on the MNIST dataset. The proposed hybrid CNN-LSTM structure has been compared to conventional intelligent machine learning methods, and the results demonstrate the superior performance of the hybrid CNN-LSTM, showcasing heightened accuracy, sensitivity, specificity, and other evaluation metrics.

**Keywords:** Deep learning, feature extraction, handwriting recognition, convolutional neural network, long-short term memory.

**INTRODUCTION**

The exploration of character recognition across various languages has been extensively conducted by researchers,<sup>[1-3]</sup> with a notable emphasis on handwritten digit identification (HDI). In the realm of information manipulation, HDI assumes a pivotal role, particularly considering the abundance of data in the form of printed text or images.<sup>[4]</sup> Compared to manual processing of printed material, the analysis of digital information proves to be a more cost-effective approach. The primary goal of HDI algorithm is to identify and translate handwritten digits into machine-readable formats. Nowadays, HDI has garnered significant attention within the research community due to its diverse applications.<sup>[5,6]</sup> HDI framework exhibits the capability to decipher content on printed pages, enabling researchers to uncover valuable information stored in historical documents and files that may be imperceptible to the human eye.

In addition, HDI techniques play a crucial role in the digital transformation of businesses and institutions. Automated HDI system contributes to various domains, such as the automatic detection of vehicle number plates,<sup>[7]</sup> recognition of digits on medical receipts to assist chemists,<sup>[8]</sup> and aiding psychologists in analyzing patient personalities.<sup>[9]</sup> However, given the substantial databases in the aforementioned areas, there is a pressing need for automated HDI system to be efficient,

robust, and capable of delivering consistent results within a minimal execution time.

The conventional intelligent approaches including support vector machine (SVM),<sup>[10,11]</sup> random forest,<sup>[12]</sup> decision tree,<sup>[13]</sup> etc., have shown credibility for HDI; however, they still have limitations in handling the inherent diversity present in individual handwriting styles. The small differences in strokes of pen and letter shapes make handwriting identification difficult and inaccurate. In addition, these methods have challenges in accurately representing complex functions and processing large datasets due to the learning capabilities. The efficacy of these approaches in classification and recognition is affected by the extraction of features from the raw data. Thus, there is a

**Corresponding Author:**

Fatimatelbatoul M. Husari, Department of Informatics and Software Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq.

E-mail: fatimatelbatoul.mahmoud@cihanuniversity.edu.iq

**Received:** June 06, 2024

**Accepted:** September 12, 2024

**Published:** October 01, 2024

**DOI:** 10.24086/cuesj.v8n2y2024.pp99-103

Copyright © 2024 Fatimatelbatoul M. Husari, Mohammed A. Assaad. This is an open-access article distributed under the Creative Commons Attribution License.

need for automatic feature learning directly from the collected data. To overcome these disadvantages, the development of advanced models capable of acquiring hierarchical features and capturing the temporal dependencies naturally present in handwritten strokes is necessary.

In this work, the novel hybrid deep learning models have been developed and employed for handwritten digit identification, namely two-dimension convolution neural network and long-short-term memory (LSTM) model. This hybrid structure has the capability to automatically acquire features, resulting in an improved data representation for handwritten recognition. The convolutional neural network (CNN) model employs convolutional and pooling layers, for learning representations directly from raw data. LSTM models with temporal memory, designed for sequential data, demonstrate proficiency in temporal information encoding. The integration of these architectures establishes a robust framework that effectively addresses the challenges of current methods, resulting in improved identification performance. In addition, the combination of 2DCNN and LSTM structures aims to leverage the strengths of both architectures to enhance the performance of handwritten recognition. In this work, 2DCNN has been used for automatic extraction of abstract features (spatial features) from the image dataset by stacking multiple layers of convolution and max-pooling, whereas LSTM is used for capturing the temporal features and effectively manage the images sequence due to the capabilities of long-term memory.

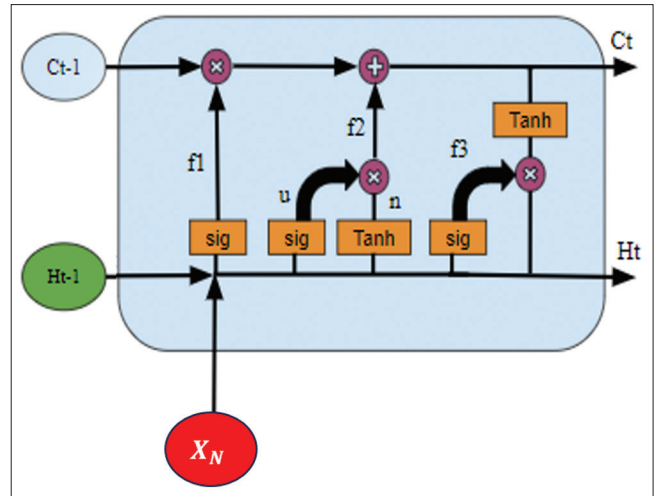
This paper is divided into seven sections including this section. Part II describes the proposed hybrid CNN-LSTM structure. The comprehension evaluation findings and discussions are presented in part III. Part V summarizes the whole work.

## A DETAILED ANALYSIS OF THE PROPOSED HYBRID CNN-LSTM ARCHITECTURE

### The Analytical Description of the Proposed Algorithm

CNN stands out among various deep learning techniques and has demonstrated significant success in intelligent classification.<sup>[14]</sup> In addition, LSTM is introduced to address the limitations of recurrent neural networks,<sup>[15,16]</sup> particularly in handling long-term dependencies within data, making it a suitable solution for mitigating the vanishing and exploding gradient problem. The LSTM architecture is illustrated in Figure 1. LSTM is basically used in processing sequential dataset and has achieved remarkable performance in various applications such as translation, natural language processing, and sentiment analysis. For MNIST image dataset, each image can be represented as a sequence of pixels. LSTM is used here to capture the sequential nature of the image pixels and recognize the pattern. Hence, combining CNN and LSTM into one single structure is a logical choice to leverage the advantages of both in handwritten recognition.

For a more detailed analysis and explanation, the mathematical analysis of the hybrid CNN-LSTM methodology can be described precisely as the following:<sup>[17]</sup>



**Figure 1:** Basic structure of long-short-term memory model

The mathematical equation of the convolution operation can be presented as:

$$\text{Conv}\left(\text{In}^{[l-1]}, F^{(N)}\right)_{X,Y} = \sum_{k=1}^{M_h^{[l-1]}} \sum_{n=1}^{M_w^{[l-1]}} \sum_{j=1}^{M_c^{[l-1]}} F_{k,n,j}^{(N)}$$

$$\text{In}^{[l-1]}_{X+k-1,Y+n-1,j}$$

$$\forall N \in [1, 2, \dots, M_c^{[l]}]$$

$$\text{shape}\left(\text{Conv}\left(\text{In}^{[l-1]}, F^{(N)}\right)\right) = (M_h^{[l]}, M_w^{[l]}) \quad (1)$$

Here  $\text{In}^{[l-1]}$  denotes the input with shape  $(M_h^{[l-1]}, M_w^{[l-1]}, M_c^{[l-1]})$  for  $[l-1]$  layer. Where  $\text{In}^{[0]}$  the input image.

For  $l$  input layer, the convolution operation between  $\text{In}^{[l]}$  and the filter/kernel  $F^{(N)}$  with size  $M_c$  filter number can be described as follows:

$$\text{In}^{[l]} = \text{Conv}(\text{In}^{[l-1]}, F^{(1)}), \text{Conv}(\text{In}^{[l-1]}, F^{(2)}), \dots,$$

$$\text{Conv}(\text{In}^{[l-1]}, F^{(M_c^{[l]})})$$

$$\text{shape}(\text{In}^{[l]}) = (M_h^{[l]}, M_w^{[l]}, M_c^{[l]}) \quad (2)$$

$M_h^{[l]}, M_w^{[l]}, M_c^{[l]}$  denote the height, width, and channel number of the image at  $l^{\text{th}}$  layer, respectively.

Further, the ReLU activation function is followed:

$$\text{In}^{[l]} = \text{ReLU}(\text{In}^{[l]}) \quad (3)$$

The output shape will remain the same as in Eq (2). After that, the maxpooling layer is used for reducing the feature's dimensionality, which in turn speed up the training process. Maxpooling operation can be expressed as:

$$\text{Pooling}(\text{In}^{[l]})_{X,Y,Z} = \max((\text{In}^{[l]})_{X+k-1,Y+n-1,Z},_{k,n \in [1,2,\dots,p[l])}] \quad (4)$$

Where  $p^{[l]}$  is the size of pooling window for  $l^{\text{th}}$  layer. The output of final layer is converted to one-dimensional vector, presented as  $(X_N^{[1]})$  with  $N$  neuron. This  $X_N$  vector represents the main features extracted by CNN layer and is directly inputted into LSTM layer as described in the following analysis equation:

The initial step is deciding which information will be excluded from the previous cell state  $C_{t-1}$  using forget gate  $f_1$ . The forget gate applies sigmoid function to the previous output  $H_{t-1}$  and the feature vector  $X_N$  to change the learning parameters weights  $W_f$  and bias  $b_f$  of forget gate.

$$f_1 = sig(W_f [H_{t-1}, X_N] + b_f) \tag{5}$$

After that, the new data can be appended to the current cell state  $C_t$  using input gate  $f_2$ , where  $f_2$  is multiplication  $u$  (using sigmoid to update the values) and  $n$  (using tanh to create new values)

$$u = sig(W_u [H_{t-1}, X_N] + b_u) \tag{6}$$

$$n = tanh(W_n [H_{t-1}, X_N] + b_n) \tag{7}$$

$$f_2 = u * n$$

Here,  $W_u$ ,  $W_n$ ,  $b_u$ , and  $b_n$  are the weights and biases, respectively.

Then, the new and updated information are added to the new current cell state as follows:

$$C_t = f_1 * C_{t-1} + f_2 \tag{8}$$

The final step is to decide which part of cell state will be the final output  $H_t$  using sigmoid function and tanh function as:

$$f_3 = sig(W_o [H_{t-1}, X_N] + b_o)$$

$$H_t = f_3 * tanh(C_t) \tag{9}$$

### Hybrid 2DCNN-LSTM framework

Figure 2 illustrates the hybrid 2DCNN-LSTM model. This model is specifically crafted for handwritten recognition and identification. Leveraging the well-established LeNet-5 structure, a widely employed CNN architecture is known for its noteworthy identification and classification capabilities.<sup>[18]</sup>

The proposed framework is structured as follows: Initially, CNN is employed to extract meaningful features from the MNIST images dataset for handwritten digit recognition. Then, the extracted features are inputted into LSTM to capture long-term dependencies in the dataset to memorize the pattern. Ultimately, the Softmax classifier is utilized to identify handwritten digits.

The primary CNN network includes three convolutional and max pooling layers, one LSTM layer with 128 neurons and one dense layer with Softmax. The first convolutional layer comprises 8 filters with a size of  $3 \times 3$ . Following this, a max

pooling layer with a size of  $2 \times 2$  is introduced by taking the maximum value from the window of pooling. Subsequently, the filter size remains constant while the number of filters is twofold. After extracting the spatial features from CNN, they are inputted into the LSTM layer to process temporal features and recognize the pattern. In the end, the Softmax classifier is used to identify the handwritten digits. The configuration of CNN-LSTM is described in Table 1.

## RESULTS AND OBSERVATIONS

The hybrid CNN-LSTM framework was examined and evaluated using the MNIST dataset which consists of 70,000 images of handwritten digits ranging from 0 to 9 with a shape of  $28 \times 28$ . 70% of the MNIST dataset is allocated for training the hybrid CNN-LSTM and the existing machine learning models, and 30% of the MNIST dataset was intended to assess the performance of the trained networks to observe how good the model is classifying the output or predicting the behavior of the model based on the given dataset. In addition, the effectiveness and performance of the hybrid deep learning structures and state-of-the-art machine learning models have been evaluated and compared using different evaluation metrics such as classification accuracy, confusion matrix, and F1-score. The outcomes of classification accuracy and F1-score for the hybrid deep learning architectures, namely CNN-LSTM and the state-of-the-art machine learning methods (SVM, decision tree, and random forest), are illustrated in Figures 3 and 4, respectively. The results have proved that the performance of the hybrid CNN-LSTM (98% accuracy) outperforms the conventional machine learning models; SVM (94% accuracy), random forest (89% accuracy), and decision tree (74%). The higher accuracy of the hybrid CNN-LSTM framework reveals the capability of CNN in capturing useful information and identifying the significant patterns/trends from the raw data and LSTM network which has ability to memorize the similar notable features of one class in the dataset.

The superior F1-score of hybrid CNN-LSTM (98%) has been noticed compared to the other traditional machine learning methods. This demonstrates that the hybrid CNN-LSTM has both high precision and high recall. To this end, the proposed CNN-LSTM structure is robust at making accurate predictions while minimizing false predictions. This shows that the proposed CNN-LSTM is performing well in recognizing and identifying the handwritten digits.

For further evaluation, the confusion matrix has been considered. [Figures 5-8] illustrate the confusion matrices of

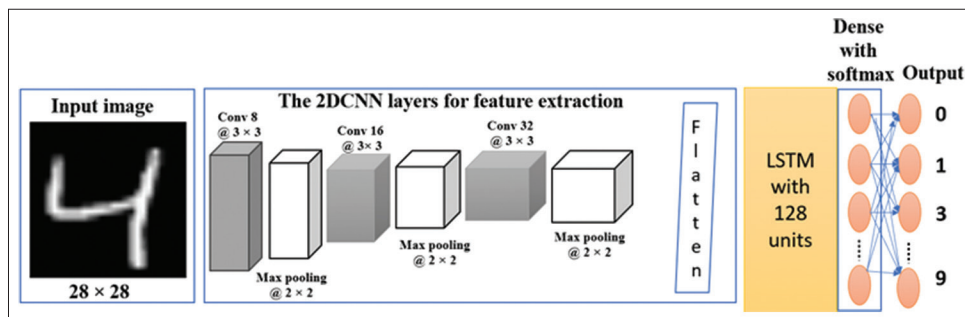


Figure 2: Proposed hybrid 2D convolutional neural network long-short-term memory framework

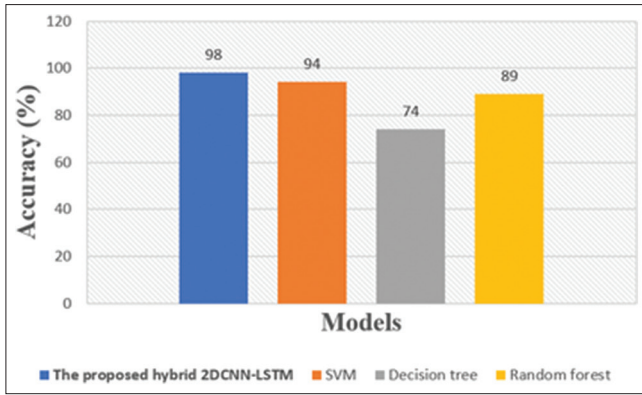


Figure 3: Overall accuracy performance for all models

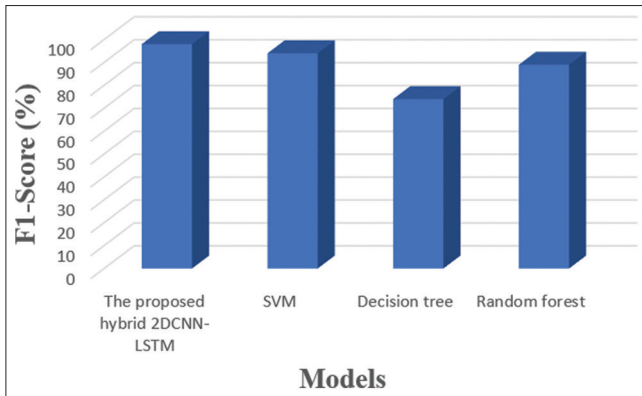


Figure 4: F1-score metric for all models

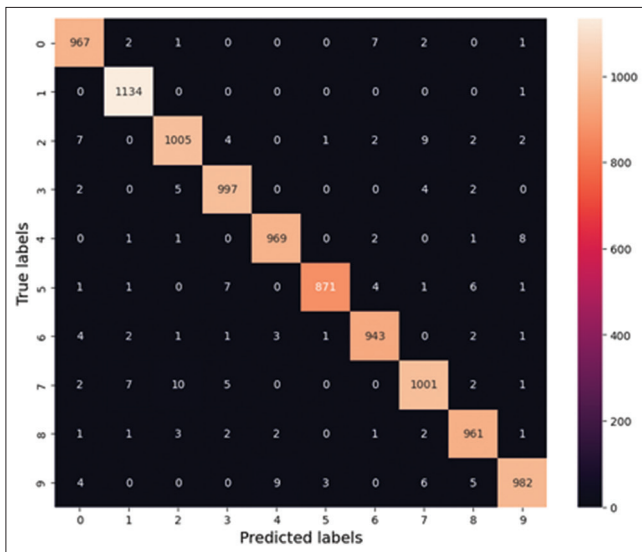


Figure 5: Confusion matrix of the proposed hybrid convolutional neural network long-short term memory model

the proposed hybrid deep learning networks and state-of-the-art machine learning models.

In these matrices, the rows represent the true labels, while the columns represent the predicted labels. It can be observed that all handwritten digit classes are well identified and recognized in the proposed hybrid CNN-LSTM compared

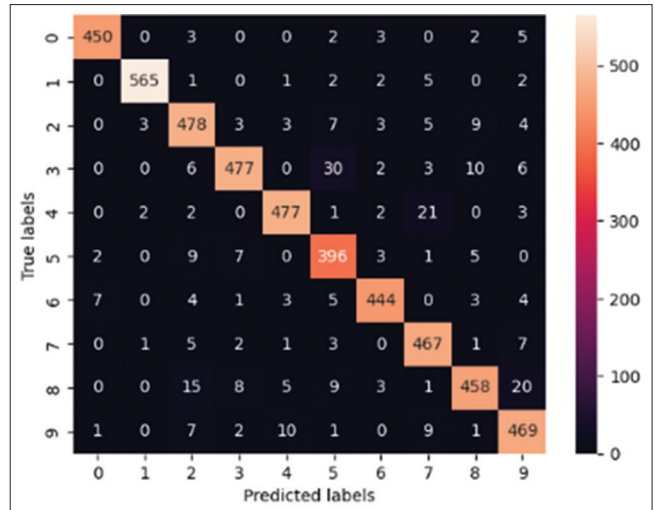


Figure 6: Confusion matrix of support vector machine model

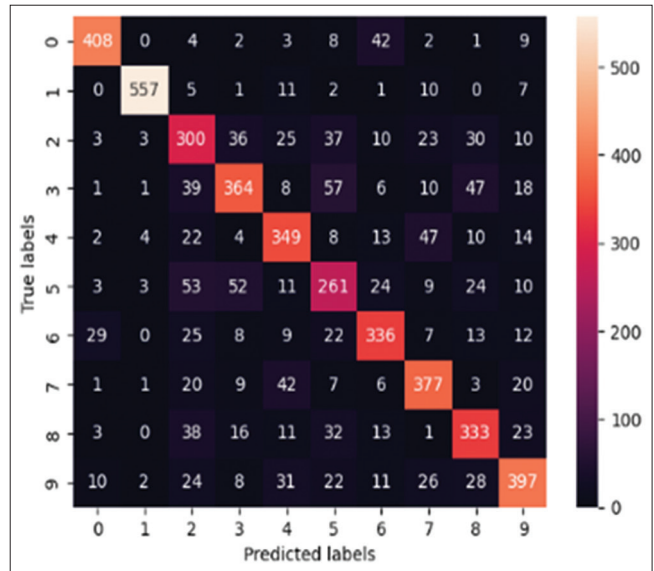


Figure 7: Confusion matrix of decision tree model

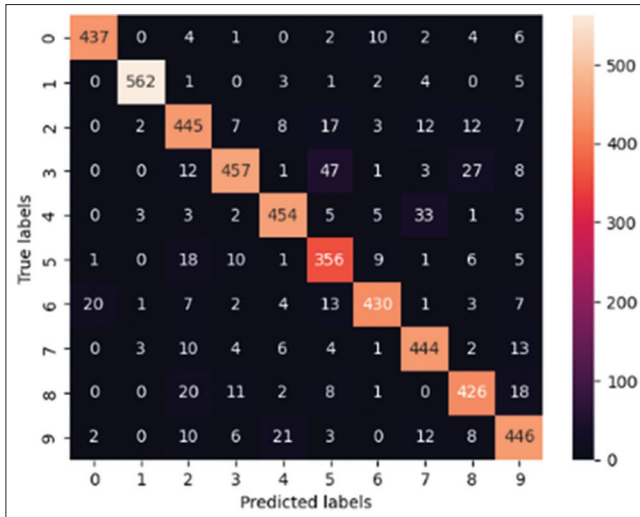
Table 1: Hybrid 2DCNN-LSTM configuration

Layer	Layer size
Convolutional 1	conv_1(28×28×8)
Max-pool 1	pool_1(14×14×8)
Convolutional 2	conv_2(14×14×16)
Max-pool 2	pool_2(7×7×16)
Convolutional 3	conv_3(7×7×32)
Max-pool 3	pool_3(3×3×32)
LSTM	lstm (128)
Dense	dense (10,1)

CNN: Convolutional neural networks, LSTM: Long-short-term memory

to the other traditional machine learning techniques. This indicates that the hybrid model precisely identifies the true prediction instances for each class of handwritten digits with a lesser number of false prediction instances.





**Figure 8:** Confusion matrix of random forest model

From an overall comparison, it can be concluded that the proposed methodology, namely CNN-LSTM, is found to be robust, efficient, and accurate for recognition and identification of handwritten digits with optimal model design and best performance.

## CONCLUSION

In this paper, a hybrid deep learning approach, namely 2DCNN-LSTM, has been developed for handwritten digits recognition. The proposed CNN-LSTM framework has evaluated and compared to the other conventional machine learning techniques to test which one would be more effective and convenient for handwritten digits classification. The hybrid CNN-LSTM architecture combines the benefits of CNN for extracting the significant features, and LSTM to process the extracted features sequentially. The outcomes proved that the learned meaningful features by CNN network are efficient and superior in accuracy 98% accuracy compared to SVM, decision tree, and random forest (94%, 74%, and 89%), respectively. This demonstrates the capability to combine the merits of the CNN network that handles the raw data by storing the trends, while LSTM network memorizes similar features within a single class in the dataset. In short, the outcomes prove that the proposed hybrid CNN-LSTM structure is efficient and highly accurate in recognizing and identifying handwritten digits.

## REFERENCES

1. B. Vinush, V. Indhuja and S. Reddy. Advancing Optical Character Recognition for Handwritten Text: Enhancing Efficiency and Streamlining Document Management. In: *2023 14<sup>th</sup> International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, pp. 1-8, 2023.
2. T. Yamasaki and A. Hiramatsu. A Study of Correcting Handwritten Answers for Short Essay Self-learning Systems. In: *2023 14<sup>th</sup> IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Koriyama, Japan, pp. 241-246, 2023.
3. P. P. Jain and H. R. Mamatha. Intelligent Recognition of Ancient Brahmi Characters Using Transfer Learning. In: *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India, pp. 1-6, 2023.
4. S. Surana, K. Pathak, M. Gagnani, V. Shrivastava, M. T. R. and G. S. Madhuri. Text Extraction and Detection from Images Using Machine Learning Techniques: A Research Review. In: *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, Tuticorin, India, pp. 1201-1207, 2022.
5. P. R. Bagde and A. A. Gurjar. A Handwritten Recognition for Free Style Marathi Script Using Genetic Algorithm. In: *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, Jalgaon, India, pp. 43-48, 2016.
6. S. Gotovac, Ž. Marušić, D. Zelenika and Ž. Šeremet. Bimodal System for Recognition of Authors of Handwritten Documents. In: *2015 4<sup>th</sup> Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro, pp. 203-207, 2015.
7. N. Shelke, S. Jadhav, M. Doifode, Y. Umate, R. Patil and N. Harinkhede. A Review on Identification of Number Plate and Wrong Way Vehicles Detection. In: *2023 IEEE International Student' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, pp. 1-4, 2023.
8. T. P. Hong, W. C. Chen, C. H. Wu, B. W. Xiao, B. Y. Chiang and Z. X. Shen. Information Extraction and Analysis on Certificates and Medical Receipts. In: *2022 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, pp. 1-4, 2022.
9. D. Redelmeier, N. Umberin and E. Etchells. Understanding patient personality in medical care: Five-factor model. *Journal of General Internal Medicine*, vol. 36, pp. 2111-2114, 2021.
10. W. Liu, J. Wei and Q. Meng. Comparisons on KNN, SVM, BP and the CNN for Handwritten Digit Recognition. In: *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, pp. 587-590, 2020.
11. A. Mahmoodzadeh, S. Rashidi, A. H. Mohammed, H. F. H. Ali and H. H. Ibrahim. Machine Learning Approaches to Enable Resource Forecasting Process of Road Tunnels Construction. In: *4<sup>th</sup> International Conference on Communication Engineering and Computer Science (CIC-COCOS'22)*.
12. K. Lavanya, S. Bajaj, P. Tank and S. Jain. Handwritten Digit Recognition Using Hoeffding Tree, Decision Tree and Random Forests - A Comparative Approach. In: *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, India, pp. 1-6, 2017.
13. Z. Li, L. Ma, X. Ke and Y. Wang. Handwritten Digit Recognition Via Active Belief Decision Trees. In: *2016 35<sup>th</sup> Chinese Control Conference (CCC)*, Chengdu, China, pp. 7021-7025, 2016.
14. A. Saffari, M. Khishe, M. Mohammadi, H. Mohammed and S. Rashidi. DCNN-fuzzyWOA: Artificial intelligence solution for automatic detection of covid-19 using X-ray images. *Computational Intelligence and Neuroscience*, vol. 2022, p. 5677961, 2022.
15. M. Assaad, R. Boné and H. Cardot. A new boosting algorithm for improved time-series forecasting with recurrent neural networks. *Information Fusion*, vol. 9, pp. 41-55, 2008.
16. A. E. Francis. Time Processing of Brain's Electrical Signals Using Normalized Constant-Correlation Function to Diagnose Brain Lesions. In: *Cihan University-Erbil Conferences, 3<sup>rd</sup> International Conference on Communication Engineering and Computer Science (CIC-COCOS'19)*, 2019.
17. F. Husari and J. Seshadrinath. Stator turn fault diagnosis and severity assessment in converter-fed induction motor using flat diagnosis structure based on deep learning approach. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 11, no. 6, pp. 5649-5657, 2023.
18. G. A. Mulla and Demir, Y. The use of clustering and classification methods in machine learning and comparison of some algorithms of the methods. *Cihan University-Erbil Scientific Journal*, 7(1), 52-59, 2023.