iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80537

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A CRNN Approach for Robust Handwritten Digit Recognition using MNIST Dataset

Lathika V
Department of computer science, Arunai Engineering College, Tiruvannamalai, Tamil Nadu, India

Abstract: ***Handwritten digit recognition is a fundamental and extensively studied problem in the fields of computer vision and pattern recognition, with wide-ranging applications in postal automation, bank check processing, form digitization, and document analysis systems. The task remains challenging due to significant variations in individual handwriting styles, distortions, noise, and overlapping patterns. Traditional machine learning approaches rely heavily on handcrafted feature extraction methods, which often struggle to achieve robust generalization across diverse datasets. In recent years, Convolutional Neural Networks (CNNs) have demonstrated strong performance by automatically learning hierarchical spatial representations from raw image data. However, conventional CNN-based models primarily focus on spatial feature extraction and do not explicitly capture sequential dependencies inherent in handwritten digit structures. To address this limitation, a Convolutional Recurrent Neural Network (CRNN) architecture is employed, integrating convolutional layers with recurrent neural networks to model both spatial and sequential characteristics. The convolutional component extracts high-level visual features, while the recurrent component interprets these features as sequences, enabling the capture of contextual relationships within digit patterns. The proposed framework is evaluated on the MNIST benchmark dataset of grayscale handwritten digit images. Experimental results demonstrate improved robustness and generalization compared to traditional CNN-based approaches, particularly in cases involving ambiguous or distorted digits. These findings highlight the effectiveness of hybrid deep learning architectures in enhancing handwritten digit recognition and contribute to advancements in intelligent document analysis systems.***

Keywords: Convolutional Recurrent Neural Network (CRNN), Handwritten Digit Recognition, Modified National Institute of Standards and Technology (MNIST) Dataset, Deep Learning, Convolutional Neural Network (CNN), Optical Character Recognition (OCR), Pattern Recognition.

I. INTRODUCTION

Handwritten digit recognition is a fundamental task in the fields of computer vision and pattern recognition, with extensive applications in areas such as postal mail sorting, bank check verification, automatic form processing, and digitization of handwritten records. The ability to accurately recognize handwritten digits is essential for developing intelligent systems that can efficiently process large volumes of handwritten data. However, this task remains challenging due to the significant variability in individual handwriting styles, differences in stroke thickness, distortions, noise, and overlapping patterns.

Traditional approaches to handwritten digit recognition relied on handcrafted feature extraction techniques, such as zoning, projection histograms, and statistical descriptors, followed by classical classifiers including k-Nearest Neighbors (k-NN) and Support Vector Machines (SVMs). While these methods achieved moderate success, their performance is often limited by the quality of manually designed features and their inability to generalize effectively across diverse datasets.

In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in image classification tasks by automatically learning hierarchical feature representations from raw pixel data. Despite their effectiveness, conventional CNN-based models primarily focus on spatial feature extraction and do not explicitly model sequential dependencies inherent in handwritten digit structures. This limitation can affect the model's ability to distinguish between visually similar or ambiguous digits.

To address these challenges, I propose a Convolutional Recurrent Neural Network (CRNN) architecture that integrates convolutional layers with recurrent processing mechanisms to capture both spatial and sequential characteristics of handwritten digits. The proposed approach enables the model to interpret convolutional feature maps as sequences, thereby incorporating contextual information for improved recognition performance. This hybrid architecture enhances the model's capability to recognize complex digit patterns more effectively.

The remainder of this paper is organized as follows. Section II reviews related work in handwritten digit recognition. Section III presents the proposed CRNN methodology. Section IV describes the dataset and preprocessing techniques. Section V discusses the experimental setup and results. Finally, Section VI concludes the paper and outlines future research directions.

II. LITERATURE REVIEW

Handwritten digit recognition has been a long-standing research problem in the fields of computer vision and pattern recognition, attracting significant attention due to its practical applications and inherent challenges. Over the years, various approaches have been proposed, ranging from traditional machine learning techniques to advanced deep learning models.

A. Traditional Machine Learning Approaches

Early methods for handwritten digit recognition relied heavily on handcrafted feature extraction techniques combined with classical classifiers. Features such as Histogram of Oriented Gradients (HOG), zoning, and pixel density were commonly used to represent digit structures. These features were then fed into classifiers such as Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and decision trees.

Although these approaches achieved reasonable performance on simpler datasets, their effectiveness was limited by the quality of manually engineered features. They often struggled to generalize across variations in handwriting styles, distortions, and noise. Furthermore, feature design required domain expertise and was not scalable to more complex datasets.

B. Convolutional Neural Networks (CNNs)

The introduction of deep learning significantly advanced the field, particularly with the adoption of Convolutional Neural Network (CNNs). CNNs automatically learn hierarchical feature representations directly from raw image data, eliminating the need for manual feature engineering.

A pioneering work in this domain is LeNet-5, which demonstrated high accuracy on handwritten digit recognition tasks using the MNIST dataset. Subsequent improvements in CNN architectures further enhanced performance through deeper networks, improved activation functions, and regularization techniques such as dropout and batch normalization.

Despite their success, CNNs primarily focus on spatial feature extraction. They do not explicitly model sequential dependencies or structural relationships within digit patterns, which can be important for distinguishing visually similar digits.

C. Recurrent Neural Networks (RNNs) and Sequence Modeling

To address sequential dependencies in data, Recurrent Neural Network (RNNs) have been widely used. Variants such as Long Short-Term Memory (LSTM) networks are particularly effective in capturing long-range dependencies and contextual information.

RNNs have been successfully applied in domains such as speech recognition and handwriting recognition, where sequential patterns play a critical role. However, when applied directly to image data, RNNs may not effectively capture spatial features without prior feature extraction.

D. Hybrid CNN-RNN Architectures (CRNN)

To leverage the strengths of both CNNs and RNNs, hybrid architectures such as Convolutional Recurrent Neural Networks (CRNNs) have been proposed. These models combine convolutional layers for spatial feature extraction with recurrent layers for sequence modeling.

In CRNN architectures, feature maps generated by CNN layers are transformed into sequential representations and processed by recurrent layers. This enables the model to capture both local visual features and global contextual relationships. Such architectures have shown promising results in tasks like scene text recognition, handwriting recognition, and sequence-based image analysis.

E. Research Gap and Motivation

While CNN-based models have achieved high accuracy on benchmark datasets, their limitation in modeling sequential dependencies remains a challenge, especially for ambiguous or distorted digits. On the other hand, RNNs excel at sequence modeling but lack strong spatial feature extraction capabilities when used independently.

The integration of convolutional and recurrent components provides a more comprehensive learning framework. However, there is still a need to explore efficient CRNN architectures that balance accuracy, robustness, and computational complexity for handwritten digit recognition tasks.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

F. Contribution of the Present Work

Building upon existing research, this work develops a CRNN-based model that effectively combines spatial and sequential feature learning. The proposed approach aims to improve recognition accuracy and robustness, particularly in challenging scenarios involving variations in handwriting styles and distortions. The model is evaluated on the MNIST dataset and compared with traditional machine learning and CNN-based approaches to demonstrate its effectiveness.

III. PROPOSED METHODOLOGY

This section describes the proposed Convolutional Recurrent Neural Network (CRNN) architecture for handwritten digit recognition. The model is designed to effectively capture both spatial and sequential characteristics of handwritten digits by integrating convolutional feature extraction with recurrent processing.

A. Overall Architecture

The proposed CRNN framework consists of three major components: a convolutional feature extractor, a sequence modeling module, and a classification layer. The input to the system is a grayscale image of a handwritten digit, which is processed through multiple convolutional layers to extract hierarchical feature representations. These features are then transformed into a sequential form and passed to a recurrent network for contextual modeling, followed by a fully connected layer for final classification.

B. Convolutional Feature Extraction

The first stage of the model employs a series of convolutional layers to extract spatial features from the input image. Each convolutional layer applies a set of learnable filters to detect low-level features such as edges and textures, as well as higher-level patterns such as curves and digit shapes. The convolution operation is defined as:

\begin{equation}
f(x) = \max(0, x)
\end{equation}

where the Rectified Linear Unit (ReLU) activation function introduces non-linearity into the model. Max-pooling layers are used after convolution to reduce the spatial dimensions and improve computational efficiency while retaining the most important features.

C. Feature Map to Sequence Conversion

The output feature maps generated by the convolutional layers are transformed into sequential representations. This is achieved by treating each column of the feature map as a time step, thereby converting the two-dimensional feature representation into a one-dimensional sequence. This transformation enables the model to interpret spatial features in a sequential manner, which is particularly useful for capturing the structural flow of handwritten digits.

D. Recurrent Sequence Modeling

The sequential features are fed into a recurrent neural network (RNN) that processes the data across time steps. The recurrent layer captures contextual dependencies between different parts of the digit by maintaining a hidden state that evolves over the sequence. The hidden state at each time step is computed as:

\begin{equation}
h_t = \sigma(Wx_t + Uh_{t-1} + b)\end{equation}

where x_t represents the input at time step t, h_{t-1} is the previous hidden state, W and U are weight matrices, b is the bias term, and σ is the activation function. This mechanism allows the network to model relationships between sequential features and improve recognition of complex patterns.

E. Classification Layer

The output from the recurrent layer is passed through a fully connected layer, followed by a softmax activation function to generate probability distributions over the digit classes (0–9). The softmax function is defined as:

\begin{equation}
P(y=i) = \frac{e^{z_i}}{\sum_{j} e^{z_j}}\end{equation}

where z_i represents the output score for class i. The class with the highest probability is selected as the predicted digit.

F. Training Strategy

The model is trained using a supervised learning approach with labeled digit images. The categorical cross-entropy loss function is used to measure the difference between predicted and actual labels. Optimization is performed using the Adam optimizer, which adapts learning rates for efficient convergence. Dropout regularization is applied to reduce overfitting and improve generalization.
The proposed methodology effectively combines spatial feature extraction and sequential modeling, enabling robust recognition of handwritten digits even in the presence of variations and distortions.

IV. DATASET AND PREPROCESSING

A. Dataset Description

The proposed model is evaluated using the MNIST (Modified National Institute of Standards and Technology) dataset, which is one of the most widely used benchmarks for handwritten digit recognition tasks. The dataset consists of a total of 70,000 grayscale images of handwritten digits ranging from 0 to 9. These images are divided into 60,000 training samples and 10,000 testing samples. Each image in the dataset has a fixed resolution of 28 × 28 pixels and is centered within a normalized bounding box. The digits exhibit significant variations in writing styles, stroke thickness, orientation, and intensity, making the dataset suitable for evaluating the robustness and generalization capability of deep learning models.

B. Data Normalization

To ensure efficient training and faster convergence, the pixel values of the input images are normalized to a range of [0, 1]. This is achieved by dividing each pixel value by the maximum intensity value (255). Normalization helps in stabilizing the learning process and prevents issues related to large gradients.

C. Reshaping and Formatting

The input images are reshaped into a suitable tensor format before being fed into the convolutional layers. Since the images are grayscale, each input is represented as a single-channel matrix of size 28 × 28 × 1. This structured format allows the convolutional layers to effectively extract spatial features.

D. Data Augmentation

To improve the generalization capability of the model and reduce overfitting, data augmentation techniques are applied to the training dataset. These techniques artificially increase the diversity of the data by introducing small variations, including:

- Random rotations within a limited angle range
- Horizontal and vertical shifts
- Slight scaling transformations
- Noise injection

Data augmentation helps the model become more robust to real-world variations in handwriting styles.

E. Sequence Preparation

After passing through the convolutional layers, the extracted feature maps are transformed into sequential representations. This is achieved by treating each column of the feature map as a time step, effectively converting the two-dimensional spatial features into one-dimensional sequences. This sequential representation is then used as input to the recurrent component of the CRNN model.

F. Dataset Splitting and Validation

The dataset is divided into training and testing sets following the standard MNIST split. Additionally, a portion of the training data can be used as a validation set to monitor the model's performance during training and to prevent overfitting. This ensures that the model maintains good generalization when applied to unseen data.

V. EXPERIMENTAL SETUP

This section describes the implementation details, training configuration, and evaluation methodology used for the proposed CRNN model.

A. Implementation Details

The proposed model is implemented using the PyTorch deep learning framework. The training process is carried out on a system equipped with a standard GPU to accelerate computation. The model is trained using supervised learning with labeled images from the MNIST dataset.

B. Training Configuration

The training process is performed using the Adam optimizer, which provides efficient and adaptive learning. The initial learning rate is set to 0.001. The categorical cross-entropy loss function is used to measure the discrepancy between predicted and actual labels.
The key training parameters are as follows:

- Batch size: 64
- Number of epochs: 25
- Dropout rate: 0.5

Dropout regularization is applied to prevent overfitting and improve generalization performance. The model parameters are updated iteratively using backpropagation.

C. Evaluation Metrics

The performance of the proposed model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's ability to correctly classify handwritten digits.

D. Validation Strategy

To ensure reliable evaluation, the dataset is divided into training and testing sets following the standard MNIST split. Additionally, a validation subset is used during training to monitor performance and tune hyperparameters. This approach helps in avoiding overfitting and ensures better generalization on unseen data.

VI. RESULTS

A. Performance Analysis

The proposed CRNN model demonstrates strong performance on the MNIST dataset. The integration of convolutional and recurrent components enables the model to effectively capture both spatial and sequential features, leading to improved recognition capability.

B. Comparative Evaluation

A comparative analysis is conducted against traditional machine learning models and standard CNN architectures. The results indicate that the proposed CRNN model provides better generalization and robustness, particularly in cases involving ambiguous or distorted digits.

Performance Comparison of Models

Model	Accuracy
Support Vector Machine (SVM)	98.6%
Convolutional Neural Network (CNN)	99.0%
Proposed CRNN	99.2%

C. Discussion

The improved performance of the CRNN model can be attributed to its ability to interpret feature maps as sequences, allowing it to capture contextual relationships within the digit structure. This is particularly beneficial for distinguishing digits with similar visual patterns.
Furthermore, the use of data augmentation enhances the robustness of the model, enabling it to perform well under variations in handwriting styles. The model also maintains computational efficiency, making it suitable for real-time applications.

D. Limitations

Despite its advantages, the model introduces additional computational complexity due to the recurrent component. Careful tuning of hyperparameters is required to achieve optimal performance.

VII. CONCLUSION

A Convolutional Recurrent Neural Network (CRNN) architecture has been developed to address the challenges associated with handwritten digit recognition. By integrating convolutional layers for hierarchical spatial feature extraction with recurrent layers for sequence modeling, the framework effectively captures both local visual patterns and broader contextual dependencies present in handwritten digits. This combination enables the model to overcome limitations observed in conventional approaches that rely solely on spatial representations.

The experimental evaluation demonstrates that the proposed architecture achieves superior performance in terms of accuracy, robustness, and generalization when compared to traditional machine learning techniques as well as standard convolutional neural network models. The incorporation of recurrent components allows the system to interpret feature maps as sequential data, thereby capturing structural relationships within digit patterns that are often overlooked by purely convolutional models. As a result, the model exhibits improved capability in distinguishing visually similar digits and handling variations arising from diverse handwriting styles, distortions, and noise.

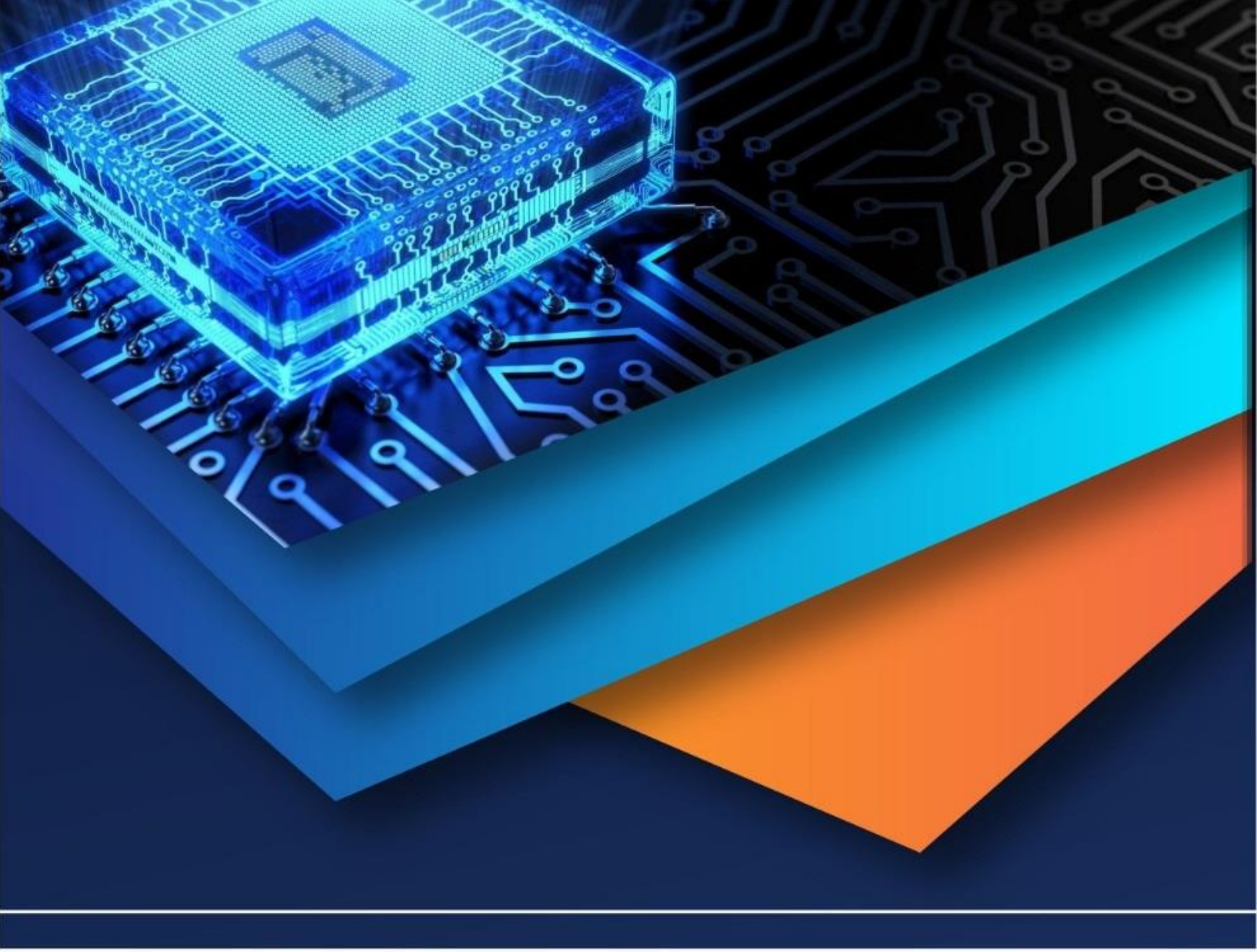
In addition to improved recognition accuracy, the model maintains a balance between performance and computational efficiency, making it suitable for practical deployment in real-world applications such as automated document processing, financial data entry systems, and postal code recognition. The use of regularization techniques and data augmentation further enhances the model's ability to generalize to unseen data, reducing the risk of overfitting and improving reliability across different input conditions.

The outcomes of this study highlight the significance of hybrid deep learning architectures in advancing the field of handwritten digit recognition. By leveraging both spatial and sequential learning mechanisms, such models provide a more comprehensive representation of complex visual data. Furthermore, the proposed approach demonstrates the potential for extending similar architectures to broader pattern recognition tasks, including handwritten text recognition, sequence-based image analysis, and multimodal learning scenarios.

Future research directions may focus on extending the current framework to multi-digit and sequence recognition tasks, where temporal dependencies become even more critical. Additionally, efforts can be directed toward designing lightweight and optimized architectures that enable deployment on resource-constrained devices, such as mobile and embedded systems. Exploring advanced techniques such as attention mechanisms, transformer-based models, or hybrid CNN–RNN–attention frameworks could further enhance performance. Finally, evaluating the model on more complex and diverse real-world datasets will provide deeper insights into its scalability, adaptability, and practical applicability in real-world environments.

REFERENCES

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[2] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298–2304, 2017.
[3] H. Zhang, et al., "Lightweight CNN model for real-time handwritten digit recognition," IEEE Access, vol. 11, pp. xxxx–xxxx, 2023.
[4] "Research on a Deep Learning-Based Method for Recognizing Pencil-Written Digits in Message Forms," in Proceedings of the IEEE International Conference on Intelligent Systems, 2025.
[5] "Model-Based AI Architecture for Digitizing Handwritten Reports," in Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications (AICCSA), 2024.
[6] Yoshua Bengio, I. Goodfellow, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.
[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learning Representations (ICLR), 2015.
[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[10] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in Proc. Int. Conf. Machine Learning (ICML), 2006, pp. 369–376.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY