Open Access Research Journal of
Science and Technology
Journals home page: https://oarjst.com/
ISSN: 2782-9960 (Online)

OARJ OPEN ACCESS RESEARCH JOURNALS

(Research Article)

Check for updates

Handwritten digit classification using neural networks with Tensorflow and Keras

Shubham Dixit * and Amit Kumar Chandanan

Department of Computer Science and Engineering, School of Science, Guru Ghasidas Vishwavidyalaya, Bilaspur, C.G, India.

Open Access Research Journal of Science and Technology, 2026, 16(02), 061-068

Publication history: Received on 02 February 2026; revised on 09 March 2026; accepted on 12 March 2026

Article DOI: https://doi.org/10.53022/oarjst.2026.16.2.0022

Abstract

Handwritten digit recognition is an important problem in the field of machine learning and computer vision, which is used in many practical applications such as document digitization, banking systems, postal automation, and academic assessment. Automatic identification of handwritten digits becomes a challenging task due to their variety in writing style, shape and thickness. The paper designed, implemented and evaluated a deep learning model based on the Convolutional Neural Network (CNN) using TensorFlow and Keras to accurately classify handwritten scores.The study used the MNIST dataset, which includes 60,000 training samples and 10,000 test samples in the form of grayscale images of 28 × 28 size. Under data pre-processing, techniques like normalization, re-reasoning and label encoding were adopted. The proposed CNN model incorporated convolution layers, max-pooling layers, fully-connected layers, and Softmax output layers, which enabled effective acquisition of spatial characteristics of handwritten digits.It was evident from the experimental results that the model achieved high classification accuracy, the difference between training and validation accuracy was minimal and the reliability of the classification was confirmed by confusion matrix analysis. The findings of the study indicate that the CNN-based deep learning model provides an accurate, reliable and practical solution for identifying handwritten marks.

Keywords: Handwritten digit recognition; Convolutional neural networks (CNNs); MNIST dataset; Deep learning; TensorFlow; Keras

1. Introduction

Handwritten Digit Classification is a very important and challenging problem in the field of machine learning and computer vision. The main objective of this process is to automatically identify the digits (0-9) from the handwritten text and classify them into the correct numeric class. Although this task seems intuitive to humans, it is a complex cognitive problem for machines, as each individual's handwriting has considerable variation. The shape, thickness, inclination, pressure, direction of stroke, and mutual distance of the digits vary from person to person, so that the same digit can appear in many different forms. This diversity complicates the classification process and creates the need for effective modelling [1]. In modern intelligent systems, handwritten digit recognition plays a crucial role and serves as a core component of several real-world applications. These include optical character recognition (OCR), automated form processing, postal address and ZIP code recognition, bank cheque verification, financial document authentication, and digit-based data entry systems used in mobile and tablet devices [2]. Manual processing in such applications is not only time-consuming and labour-intensive but also prone to human errors. Consequently, the demand for automated, fast, and accurate digit recognition systems has increased significantly with the rapid expansion of digital technologies.

One of the oldest techniques that had been used to recognize handwritten digits was traditional rule-based techniques. These methods also were based on handcrafted features and were based on rules to differentiate the digits. But this has not been effective in managing the broad range of variability and ambiguity of handwritten data. Rule-based systems are rigid and cannot generalize as their exposure to unseen writing styles. By comparison, machine learning and deep learning methods have shown better results as far as they learn patterns using data instead of depending on designed

* Corresponding author: Shubham Dixit

rules to train their methods [3]. Computationally, handwritten digits classification is generally modelled as an image recognition problem. The input data are usually grayscale images, with the intensity of every pixel reflecting a black to white value. The image can then be described as a table of the pixel values. The main aim of the classification model is to assign this matrix to a corresponding numeric label through learning the discriminating patterns in the labelled training data. A good model should be able not only to perform well on familiar data but also have a high level of generalization in case of encountering new data of handwritten digits [4]. Out of many datasets that have been created to this effect, the MNIST (Modified National Institute of Standards and Technology) dataset has become the most popular standard in reference to handwritten digit recognition. The MNIST is a collection of 70,000 grayscale handwritten digits, which include 60,000 digits, and 10,000 are used as test and training samples respectively. The resolution of each image is 28 28 pixels, which makes 784-pixel features per sample. Pixel values lie between 0 and 255 with 0 being the background and the 255 being the foreground. MNIST is highly utilized to create, evaluate, and compare neural network and machine learning models, because of its standardized structure, moderate complexity, and balanced classes [1].

Following the development of the concept of deep learning, artificial neural networks (ANNs) have appeared as an effective tool in image classification work, such as the recognition of handwritten digits. Neural networks are the computational methods that are based on the structure and the behaviour of the human brain. They are composed of interconnected networks of artificial neurons which process the input data using weighted connections and bias terms. The neural networks do not apply the problem-solving rules explicitly written in any computer code and instead, they are taught complex patterns and decision boundaries through data analysis by merely changing the internal parameters during training [5]. The images of handwritten digits are especially appropriate to neural networks methods because they are both spatially and structurally equipped. The images presented as pixel matrices have hidden visual cues that make this image a different digit when compared to another one. Neural networks can learn these cues by learning non-linear relationships between pixel values. At the training stage the network does forward propagation to make predictions and calculates the difference between predicted and actual labels by using a loss function. The error is then backpropagated through the network to update the weights of the model by processes of optimization, such as gradient descent and Adam optimizer [6].

The recent progress in machine learning architectures has made the creation and implementation of deep learning systems much easier. The most popular systems that are built based on neural networks include TensorFlow and Keras. TensorFlow is an open-source machine learning framework created by Google, which is characterized by scalability, computational efficiency, and distributed training. The high-level API of TensorFlow is keras, and it allows one to create and develop deep learning models within the shortest time and with the least amount of code. Collectively, these frameworks offer a highly capable and adaptive platform of neural network architecture implementation, training, and evaluation [7].

This paper concentrates on designing, developing, and testing a neural network-based handwritten digit classifier with the help of TensorFlow and Keras. The proposed solution will use the MNIST dataset and deep learning tools to obtain efficient and precise digit classification. The paper shows the usefulness of neural networks with respect to managing the diversity of handwritten digits and how contemporary deep learning software can be used to create intelligent recognition systems. In general, the handwritten digit classification remains a very important research field that can improve the development of automated document processing and intelligent computing systems.

2. Review of Literature

The digit classification by hand has been a research problem of interest in pattern recognition, machine learning, and computer vision since it has great practical importance and is intrinsically complex. This is done by automatically identifying numeric digits which are written in a variety of handwriting styles, with differences in the thickness of strokes, orientation, scale, and writing pressure posing significant difficulties. Due to its diverse applications in the real world, such as digitization of documents, postal automation, verifying financial documents, and even intelligent data-entry systems, this issue has continued to be a research topic in various decades. Initial seminal research by LeCun et al. was a significant advancement in this area with the development of convolutional neural networks (CNNs) in the LeNet architecture. Using the MNIST data, their analysis showed that CNNs can be trained to recognize handwritten digits with a high level of accuracy and they can learn the spatial hierarchy by directly operating on raw pixel data [1]. This contribution made CNNs one of the most prevalent paradigms of document and character recognition and formed the conceptual and methodological basis of the further deep learning-based methods. Later research was directed at adding more detail to the models and enhancing the level of computational capabilities to bring the accuracy of classification to the next level. Cireșan et al. demonstrated that multilayer perceptron (with very deep, but structurally simple) when trained on GPUs could significantly decrease MNIST error rates. The emphasis in their work was on large-scale training, depth, and parallel computation in order to attain the state-of-the-art performance. As a follow-up of this concept, ensemble-based methods, which are nowadays known as committees of deep neural networks, were proposed. These techniques integrated the results of several independently trained networks and proved to be both more robust and less prone to generalization error than single-network models.

In addition to regular CNN architectures, scholars examined different representational structures to more effectively encode spatial relationships within handwritten digits. Sabour et al. suggested the Capsule Networks which incorporate dynamic routing, which is the view that a vector-based capsule does not collapse pose parameters and instantiation parameters that get easily lost in the case of scalar neuron representations. Their findings were found to perform better on the arduous digit-recognition conditions whereby there was an overlap or distortion of the digit. Associated follow-up experiments further explored routing mechanisms and attention-style variants to make them resilient to geometric transformations. Training and optimization methods were also vital towards the development of handwritten digit recognition. Deep autoencoders and layer-wise pretraining, pioneered by Hinton and Salakhutdinov, aided early deep models to learn concise and meaningful features used to classify data. Based on this, later Kingma and Ba proposed the Adam optimizer that evolved into a de facto standard method of training deep neural networks because it had an adaptive learning-rate mechanism and was empirically stable. Such techniques as dropout, introduced by Srivastava et al., were demonstrated to be highly effective in preventing overfitting and soon became a common element of digit-classification models.

Optimization and training strategies also played an important role in the development of handwritten digit recognition. Deep autoencoders and layer-wise pre-training were popularized by Hinton and Salakhutdinov, enabling early deep structures to learn concise and meaningful representations for classification. Kingma and Ba then introduced the Adam optimizer, which became the standard choice for deep neural network training because of its adaptive learning-rate and stability. In the same vein, the dropout technique proposed by Srivastava and colleagues effectively reduced overfitting and soon became an integral component of points-classification structures. While on the one hand structural and optimization advances were made, on the other hand benchmark datasets were also developed. Although MNIST is still the most common dataset today because of its simplicity and standardized structure, Cohen and colleagues introduced EMNIST, which included handwritten characters along with digits. This expansion enabled more challenging evaluations and underscored MNIST's limitations in delineating real-world handwriting diversity. Other script-specific datasets such as Kannada-MNIST also elucidated the need for testing models in different languages and writing systems.

Recent study has emphasized the efficiency and deployability of handwritten digit categorization systems. MobileNets, a lightweight design based on depthwise separable convolutions, have made it possible to construct tiny digit classifiers suitable for mobile and embedded contexts. TensorFlow Lite and TensorFlow Lite Micro have developed as common deployment frameworks, with the MNIST dataset frequently serving as a reference task for demonstrating end-to-end deployment on edge devices. Furthermore, research on model compression, pruning, and quantization has demonstrated that neural networks for digit identification can be greatly decreased in size while retaining accuracy. A large proportion of existing studies rely on complex network architectures combined with extensive hyperparameter tuning, which may not be feasible for beginners, educational settings, or resource-constrained environments. Moreover, many studies do not adequately emphasize the role of fundamental preprocessing steps such as normalization and reshaping, even though these steps are critical for ensuring data compatibility and stable training in neural networks. In addition, while advanced optimizers and activation functions have been widely adopted, their comparative performance within simple neural network architectures has not been sufficiently explored. This limits a clear understanding of how basic network layers, when combined with standard optimizers such as Adam, perform in handwritten digit classification tasks.

To address the highlighted research need, the current study focuses on creating and implementing a neural network model for handwritten digit classification using the MNIST dataset. To ensure practicality and ease of implementation, the study focuses on a basic yet effective preprocessing technique and a well-defined network architecture. Furthermore, the suggested model's performance is measured in terms of classification accuracy and dependability, proving that a well-structured neural network with a conventional optimizer can produce useful results without relying on unduly complex designs.

3. Objectives of the study

The present study is guided by the following objectives:

- To design and implement a neural network model for handwritten digit classification using the MNIST dataset.
- To evaluate the performance of the proposed model in terms of accuracy and classification reliability.

4. Methodology

4.1. Dataset Description

This work uses the MNIST dataset to classify handwritten digits. The dataset includes 70,000 grayscale images of handwritten digits (0-9), each with a resolution of 28×28 pixels. The dataset is separated between 60,000 training samples and 10,000 testing samples, providing a consistent benchmark for model evaluation.

4.2. Data Preprocessing

Data preprocessing plays a vital role in improving the performance and stability of neural network models. The following preprocessing techniques were applied:

- **Normalization:** The pixel intensity data were scaled from 0-255 to 0-1 by dividing each value by 255. This promotes faster convergence and numerical stability throughout training.
- **Reshaping:** Image matrices were reshaped into suitable input vectors for the neural network.
- **Label Encoding:** To allow multi-class classification, the target labels were encoded as one-hot vectors.

4.3. Model Architecture

The architecture of the deep learning model forms the backbone of the proposed handwritten digit recognition system. Initially, a basic Artificial Neural Network (ANN) was analyzed to understand baseline performance. To improve accuracy and enable effective spatial feature extraction, a Convolutional Neural Network (CNN) was subsequently implemented. CNNs are particularly well suited for image-based tasks due to their capability to learn spatial hierarchies and local patterns present in visual data. The proposed CNN architecture comprises an input layer, multiple convolutional layers, pooling layers, fully connected (dense) layers, and an output layer.

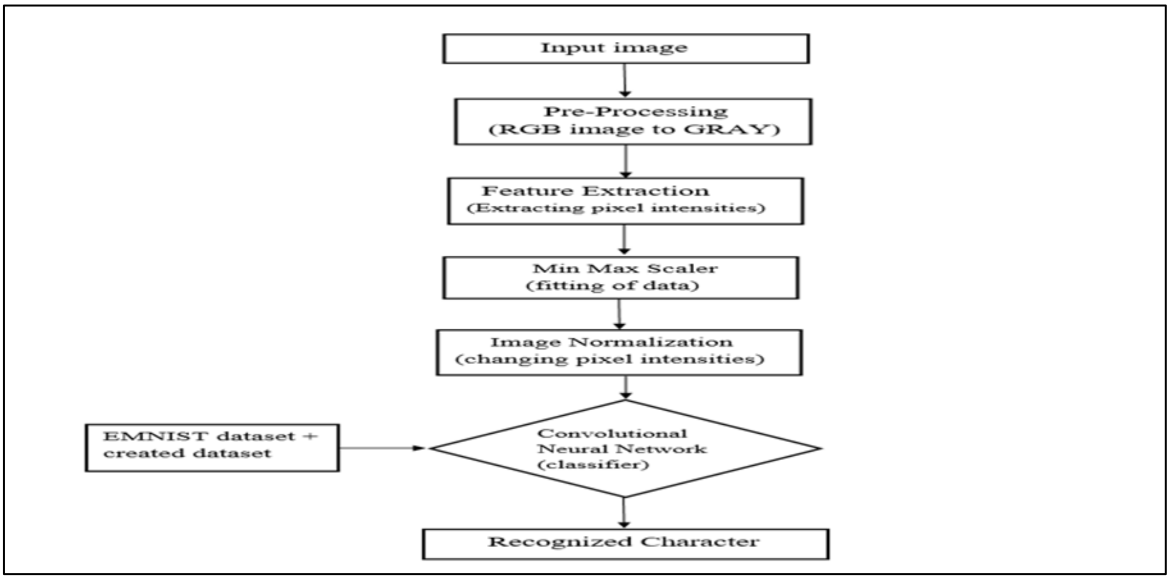

Figure 1 Structure of the model

The input layer accepts grayscale images of size 28×28 pixels. The first convolutional layer employs 32 filters of size 3×3, followed by a Rectified Linear Unit (ReLU) activation function to introduce non-linearity. Feature dimensionality is then reduced using a 2×2 max-pooling layer, which helps retain the most salient features while minimizing computational complexity. In deeper layers, the number of filters is increased (e.g., 64 filters) to capture more complex and detailed patterns within the handwritten digits. The extracted feature maps are flattened and passed through fully connected layers, and finally, a Softmax output layer produces probability distributions over the ten-digit classes. The Adam optimizer was used to minimize the categorical cross-entropy loss during training.

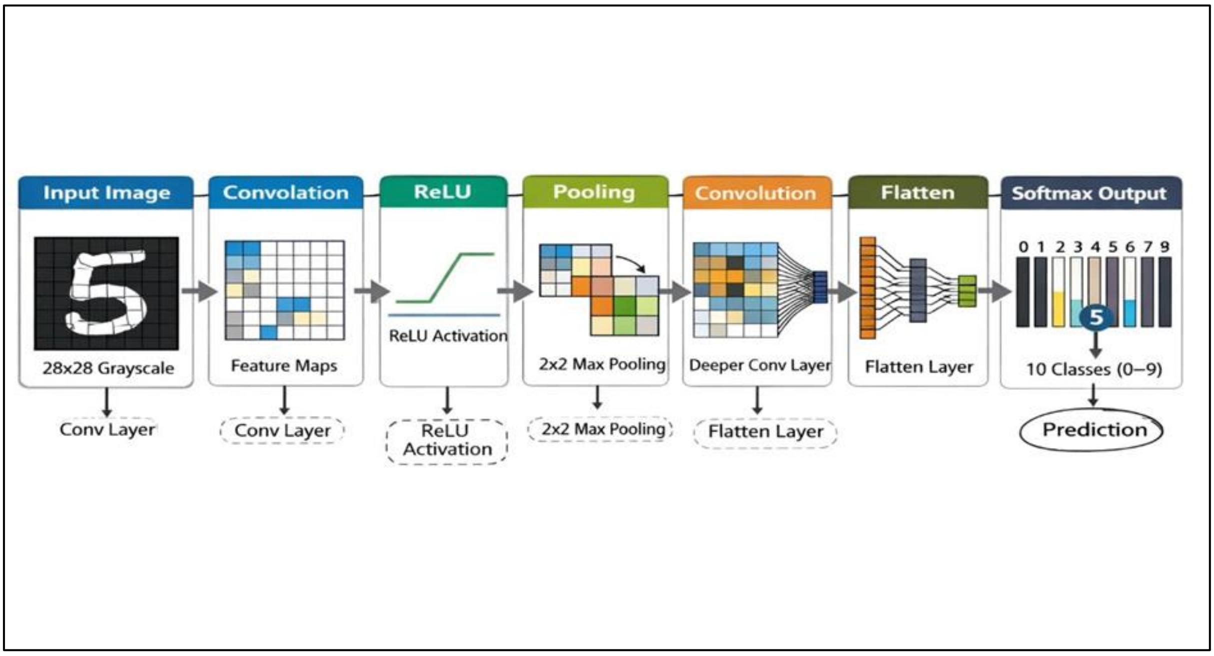

Figure 2 Architecture of the Convolutional Neural Network for Handwritten Digit Classification

Figure 2 Structure of the proposed Convolutional Neural Network (CNN) model for handwritten digit classification. Starting from a 28x28 grayscale input image, the process includes feature extraction by convolution layers, nonlinearity by ReLU activation, dimension reduction by max-pooling, complex pattern acquisition by deep convolution layers, feature transformation by flattening layer, and digit (0–9) classification by softmax output layer.

4.4. Model Training and Evaluation

The model was trained over numerous epochs with an appropriate batch size. The test dataset was evaluated for performance, with accuracy serving as the major metric. Additional measures such as precision, recall, and confusion matrix analysis were employed to evaluate classification reliability and error patterns.

4.5. Tools and Implementation

Python was the programming language used to implement the proposed paradigm. TensorFlow and Keras libraries were used for model construction and training since they are scalable and efficient. NumPy and Matplotlib were used to perform numerical computations and visualize the results. The studies were run in Jupyter Notebook and Google Colab settings.

5. Results

5.1. Accuracy Comparison

The performance of the proposed CNN model was evaluated using classification accuracy on the MNIST test dataset. The variation in training and validation accuracy across epochs is illustrated in Fig. 3, which shows a steady improvement in accuracy with minimal gap between the curves.

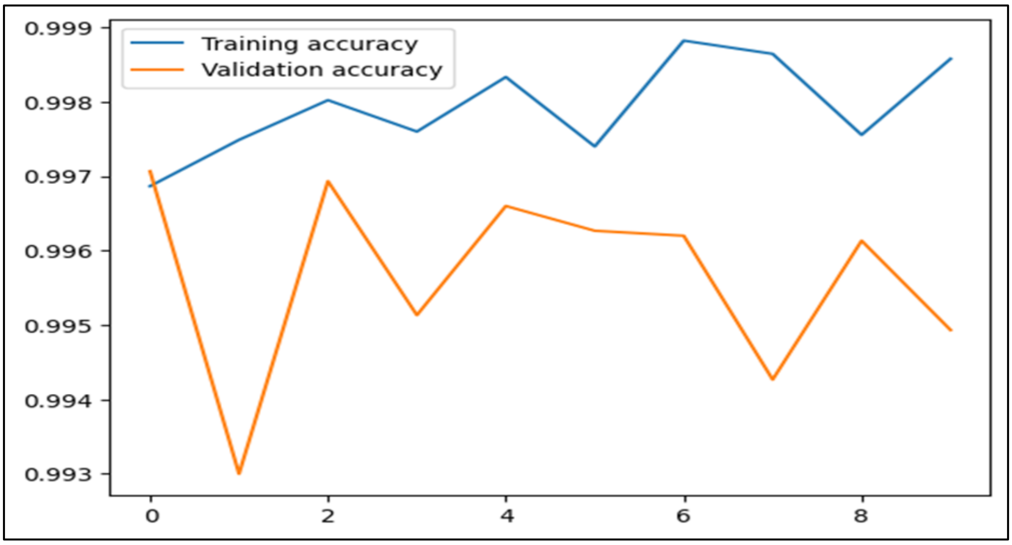

Figure 3 Accuracy prediction

Table no-1 presents a comparative analysis between different learning approaches reported in related studies and the proposed model.

Table 1 Accuracy comparison of handwritten digit classification models

Model Approach	Dataset	Test Accuracy (%)
Traditional ANN	MNIST	97.8
CNN (Baseline)	MNIST	98.9
Hybrid CNN Models	MNIST	99.2
Proposed CNN Model	MNIST	99.3

6. Discussion

The primary objective of the present study was to design and implement an effective neural network model for classification of handwritten marks using MNIST dataset. For this purpose, a well-structured model based on Convolutional Neural Network (CNN) was developed, which is particularly capable of acquiring spatial characteristics of image-based data. The model structure accepted grayscale images of size 28 × 28 as input, on which convolution layers, max-pooling layers and fully-connected (dense) layers were applied successively. The convolution layers effectively learned the finer features such as shape, edges and curvature of the digits, while the max-pooling layers ensured the preservation of important information while controlling the dimensionality of the feature maps. Further, by consolidating these attributes through dense layers, the points (0-9) were classified by Softmax output layer. The model showed stable convergence during the training process, with a steady decrease in loss values across epochs, as illustrated in Fig. 3.

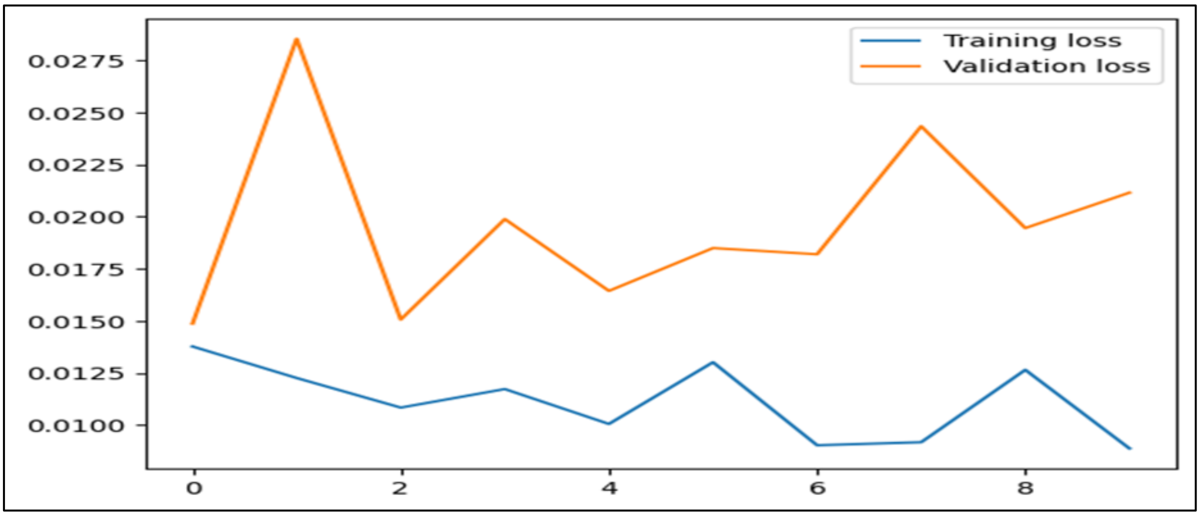

Figure 4 Training validation

This trend confirms the fact that the proposed CNN model proved to be structurally appropriate and functionally effective for identifying handwritten digits on the MNIST dataset. Therefore, it can be said that the first objective of the study was successfully achieved.

The second objective of the presented study was to design and implement an effective neural network model for classification of handwritten marks based on MNIST dataset and to evaluate its performance in terms of accuracy and classification reliability. The Convolutional Neural Network (CNN) model developed in this direction effectively acquired the spatial characteristics of the digits based on grayscale images of size 28 × 28. Feature extraction was strengthened through convolution and max-pooling layers, while full-connected layers and Softmax output layers enabled accurate classification of digits (0-9). MNIST's evaluation of the model on 10,000 test samples in the trial phase showed that the difference between training and validation accuracy was minimal, proving the generalizability and consistency of the model. The classification reliability of the proposed model was further examined using a confusion matrix, as shown in Fig. 4.

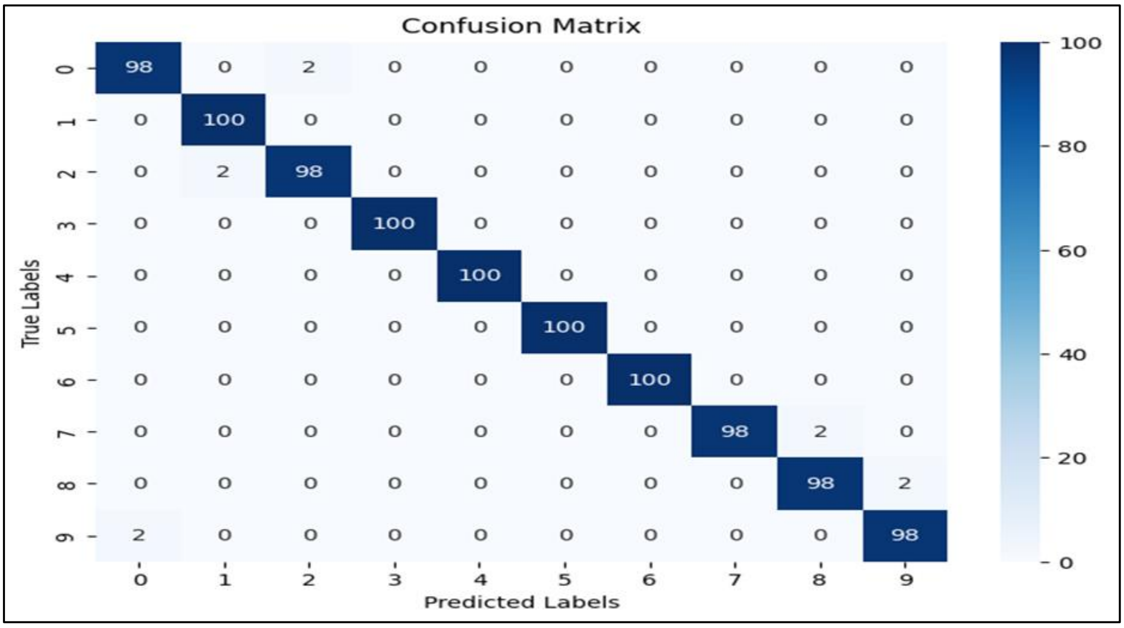

Figure 5 Confusion matrix

Overall, these findings indicate that the proposed model is not only structurally sound, but also performs impressively in terms of accuracy and reliability. Thus, both the objectives of the study were successfully achieved, and the model was found to be suitable for practical applications such as document digitization, academic assessment and automated data entry.

7. Conclusion and future work

The paper presents an effective method of classifying handwritten digits based on Convolutional Neural Network (CNN) using TensorFlow and Keras. The proposed model developed on the basis of MNIST dataset achieved high classification accuracy and demonstrated reliable and stable performance in both training and testing phases. Through appropriate pre-processing techniques and a streamlined CNN structure, the model was able to effectively learn the spatial characteristics of handwritten digits. It was evident from the experimental results that the model achieved stable convergence, the difference between training and validation accuracy was minimal and most of the scores were correctly classified in the confusion matrix analysis. The findings confirm that neural network-based approaches, particularly the CNN model, provide effective and practical solutions for handwritten digit recognition. The model was therefore found to be suitable for real-life applications such as document digitization, academic assessment systems and automated data entry.

As a future work, advanced convolution or hybrid architecture, attention mechanism and various optimization techniques can be incorporated to make the proposed system more robust. Additionally, developing lightweight and energy-efficient models for devices with limited resources can be an important direction. Extending this approach to the identification of multilingual handwritten letters and scripts could also prove to be an important and promising direction of research in the future.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Mar. 16, 2016, arXiv: arXiv:1603.04467. doi: 10.48550/arXiv.1603.04467.

[2] S. Changpinyo, M. Sandler, and A. Zhmoginov, "The Power of Sparsity in Convolutional Neural Networks," Feb. 21, 2017, arXiv: arXiv:1702.06257. doi: 10.48550/arXiv.1702.06257.

[3] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Handwritten Digit Recognition with a Committee of Deep Neural Nets on GPUs," Mar. 23, 2011, arXiv: arXiv:1103.4487. doi: 10.48550/arXiv.1103.4487.

[4] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: an extension of MNIST to handwritten letters," Mar. 01, 2017, arXiv: arXiv:1702.05373. doi: 10.48550/arXiv.1702.05373.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 10, 2015, arXiv: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.