International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

MNIST and SVHN Digit Classification

S. Chandrakala [1,*], Mariam Ghani [2], Sanath B S [2] and Swathika Murugan [2]

[1] *Department of Computer Science and Engineering SRM Institute of Science and Technology, Ramapuram, Chennai, India.*
[2] *Department of Computer Science and Business Systems, SRM Institute of Science and Technology, Ramapuram, Chennai, India.*

International Journal of Science and Research Archive, 2025, 16(01), 1919-1923

Publication history: Received on 14 June 2025; revised on 22 July 2025; accepted on 25 July 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.1.2198

Abstract

By utilising the MNIST database coupled with the SVHN data, identification of multiple handwritten digits is being achieved by the model built. In this particular use case, Convolutional neural networks (CNN) algorithm is integrated with the MNIST dataset, whereas Long short-term memory (LSTM) is employed for the SVHN dataset to sequentially classify the digits. Furthermore, the concatenation of both outputs will be trained using a final classifier. The MNIST database contains numbers ranging from 0-9, while the other database is similar in flavour, containing over 60,000 labelled images. The primary goal of this project is to develop a reliable, effective, and efficient methodology for recognizing and identifying multiple handwritten digits with minimum errors. The applications of such an accurate model lies in banking sectors, healthcare departments and many more.

Keywords: MNIST Classification; SVHN; Digit classification; Number recognition; Handwritten digits recognition; House numbers;

1. Introduction

The MNIST dataset has been the benchmark for classification of handwritten digits and the SVHN dataset provides a more complex variations of digits mimicking real world data of house numbers data gathered from google street view. The convolutional neural network (CNN) and recurrent neural network (RNN) has drastically enhanced the accuracy of digit classification. While being said existing research papers have not clearly optimised their model for deploying them in real time applications.

The proposal of a hybrid model in constructed in the basis of CNN's excellence in capturing spatial hierarchies from the images and LSTM does exceptionally well in learning sequences the classification operation can be carried out for multiple digits from an image. Our study aim is to compare effectiveness of CNN-RNN model with CNN – LSTM hybrids for digits classification. The applications of this study can be extended into works like identifying license plates, identifying numerical values in documents even if the written numerical are illegible.

2. Literature survey

Traditional machine learning models such as Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), and Multi-layer Perceptron (MLP) were initially employed for handwritten digit classification. Saeed Al-Mansoori implemented an MLP-based neural network to recognize and classify handwritten digits from 0 to 9. However, these conventional methods struggled with complex feature extraction and were not as effective in handling large datasets. Anuj Dutt demonstrated that deep learning frameworks significantly enhance classification accuracy. Using CNNs with Keras and Theano as backends, he achieved an accuracy of 98.72%. When CNN was implemented using Tensorflow, the accuracy

* Corresponding author: S. Chandrakala

improved to an outstanding 99.70%. Despite the increased computational complexity of deep learning models compared to the traditional machine learning algorithms, their superior accuracy and robustness make them the preferred choice for handwritten digit classification.

Guan Gui enhanced CNN performance by applying batch normalization and adjusting dropout rates from 0.25 to 0.4. His model achieved a validation accuracy of 96.80% while reducing overfitting through image augmentation techniques. He further demonstrated that image enhancement techniques, such as changing brightness and saturation, help improve classification accuracy while reducing validation loss. Yanling Yang and Tao Wang optimized CNN architectures using two convolutional layers and two fully connected layers to balance accuracy and computational efficiency. Agastya Gummaraju et.al compared various machine learning models, including K-NN, SVM, and deep CNN architectures such as GoogLeNet and ResNet-50 for handwritten Devangari numerals classification. Their results indicated that a well-optimized CNN model with appropriate hyperparameters outperformed even complex architectures like ResNet-50, achieving an accuracy of 96.52%.

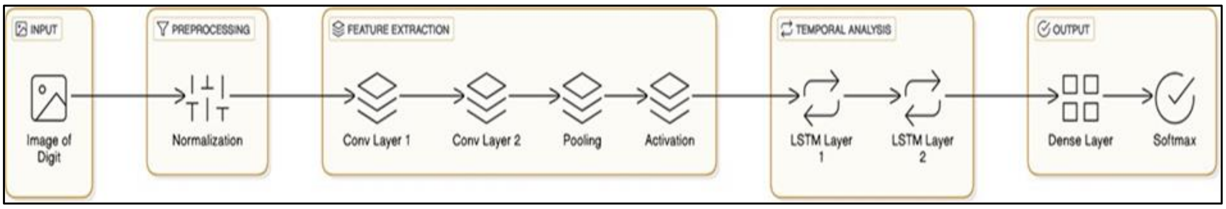

Figure 1 Flow Diagram

3. Methodology

3.1. Input layer

This research aims to utilise two datasets; MNIST, which is a collection of over 60,000 handwritten digits along with the SVHN (street view house numbers) dataset which includes over 600K labelled images of real-world house numbers from Google street View. While the handwritten digits dataset is being used for recognition of digits and being trained on neural networks, the purpose of the SVHN dataset is to learn the sequence of digits and accurately identify the numbers.

The input layer consists of these two datasets; each of them being fed to a different algorithm in order to retrieve the desired output. Furthermore, we change the range of the numbers to 0-1, instead of 0-255, as normalising the input images and grayscaling the images improves the accuracy and allows the dimensionality to be the same throughout.

3.2. Convo2D Layer

The Convolutional Neural network layers are used to extract the spatial features using 3X3 filters in order to detect patterns such as edges and curves. Moreover, it computes a two-dimensional convolutional by gliding the filter over the images and performing dot operations between the values to highlight the feature values. This is done in the initial convo 2D layer, whereas the next layer is responsible for detecting the combination of curves and edges to predict the images efficiently.

3.3. Max pooling

The max pooling layer comes after the convolutional layers and is used to enhance the accuracy of the model by enabling spatial dimensions of features while retaining the most important parts and removing unnecessary parts. Consequently, we choose a maximum value from each region after we glide the filter of 3 x 3 over the grids. This particular layer allows reduction of computational complexity and minimized overfitting, vastly improving the ability of the model to classify the digits by focusing on the relevant and integral patterns only.

3.4. Activation

The activation function being used for this research is ReLu, being one of the most common yet powerful methods for introducing non-linearity, it allows the convolutional layers to capture edges, patterns by marking the meaningful pixel activations and ignoring the weak signals. Furthermore, this creates non linearity and helps the model prevent saturation and speeds up convergence.

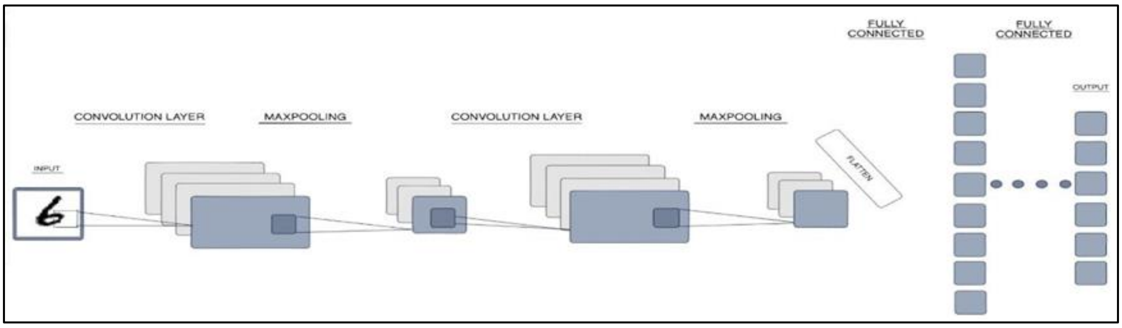

Figure 2 Convolutional Neural Network

3.5. Long short-term memory:

While the Recurrent Neural Network lays the basis for the concept of memory in neural networks, it fails to address vanishing and exploding gradients, which occur when weights of the network layers multiply multiple times before deriving the output. To address this specific problem LSTM an enhanced model of the RNN is implemented. The LSTM network functions 4 hidden layers consisting 3 sigmoid functions and 1 tanh functions.

These hidden layers form the foundation of the Forget gate, Input gate and the Output gate. The forget gate helps the model to understand the length of the sequence of the numbers which is stored in its memory mimicking a sliding window type approach except using an adaptive approach of storing numbers rather than a fixed window which does not suit the requirements as the sequence contains a range of digits between 1-7.

The input gate decides how much information is necessary to be added and the output gate to complete one sequence and the output gate decides on what is necessary to pass to the subsequent hidden state to derive predictions.

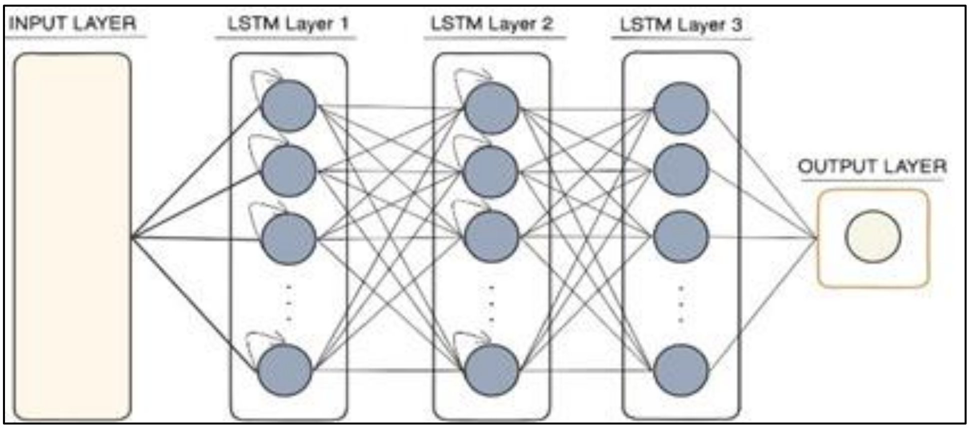

Figure 3 Long Short Term Memory

3.6. Final classifier

The final classifier is used in order to perform the final recognition of the digits and is rather an integral part of the system architecture. The output feature vectors are combined to form a comprehensive representation that encodes spatial as well as contextual information for sequencing purposes. The architecture enables the strengths of static recognition combined with sequenced learning recognition leveraging the system's capabilities and underlying applications.

4. Results and discussion

As shown in the below, the hybrid model has the capability to recognise and comprehend multiple digits by employing sequence learning patterns and is able to efficiently identify multiple digits. The hybrid model demonstrates strong performance in handwritten digit classification by integrating the spatial feature extraction capabilities of Convolutional Neural Networks (CNN) with the sequence learning power of Long Short-Term Memory (LSTM) networks. By combining these two architectures, the model effectively captures both spatial hierarchies and contextual patterns within digit representations, resulting in improved classification accuracy and robustness.

This architecture is particularly well-suited for applications requiring high precision and consistency, such as automated postal code reading, intelligent document processing, and numeric data extraction in finance and healthcare. The LSTM component enhances the model's ability to recognize temporal patterns across digit sequences, while the CNN component ensures detailed feature extraction from each input. Together, these strengths make the hybrid model a promising solution for advancing the accuracy and adaptability of digit recognition systems.

Consequently, with the usage of LSTM in the hybrid model, it reduced errors and misalignment, which further resulted in the digits being classified more adequately. The advancements made in this paper distinctly highlights the huge advantage achieved by coupling CNN with LSTM.

5. Conclusion

This research introduces a hybrid model that integrates a CNN-based MNIST classifier and an LSTM- based SVHN feature extractor, enabling sequence-level digit recognition from handwritten numerical images. CNN outperforms conventional classifiers in digit recognition by effectively extracting spatial features and hierarchical patterns from static images. Meanwhile, LSTM networks, adept at sequence modelling, enhance the recognition of multi-digit handwritten sequences by learning contextual dependencies between consecutive digits. The combination of these models ensures robust feature extraction and accurate sequence classification.

The proposed hybrid approach significantly improves recognition accuracy by leveraging CNNs for spatial learning and LSTMs for sequential understanding. The integration of both architectures enhances HDR performance, ensuring precise digit extraction from complex handwritten sequences. Future enhancements may focus on incorporating transformer-based models to further optimize sequence recognition and reduce computational overhead.

5.1. Future enhancements

This model can be catered to numerous sectors in the world, stretching from the finance sectors being used in banking to applications in the medical and healthcare industry. While these applications have not been explored and there is immense potential left to be uncovered, refinement to this research includes training the model on different handwriting styles, fonts and computer-generated texts. Furthermore, the model can be optimised by the usage of other algorithms and computational techniques in order to enhance the digit recognition process and improve the accuracy.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Saqib Ali, Zareen Sakhawat, Tariq Mahmood, Muhammad Saqlain Aslam, Zeeshan Shaukat, Sana Sahiba, A robust CNN model for handwritten digits recognition and classification. 2020.IEEE

[2] MNIST Handwritten Digit Recognition using Machine Learning. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).2022.

[3] A. Gummaraju et al. Performance Comparison of Machine Learning Models for Handwritten Devanagari Numerals Classification.

[4] Research on handwritten Digits Classification Technology Based on Convolutional Neural Network Algorithm.2023. IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI).

[5] Research on MNIST Handwritten Number Recognition Based on Convolutional Neural Networks. 2023. 2nd International Conference on Artificial Intelligence and Blockchain Technology (AIBT).

[6] Performance Analysis of a Handwritten Digit Recognition Using CNN-WOA.2023 5th International conference on Inventive Research in Computing Applications (ICIRCA).

[7] Recognition of Scanned Handwritten digits using Deep Learning.2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon)

[8] Handwritten Character and Digit Recognition with Deep Convolutional Neural Networks: A Comparative Study Chui En Mook;Chin Poo Lee;Kian Ming Lim;Jit Yan Lim 2023 11th International Conference on Information and Communication Technology (ICoICT).

[9] G. Zhu, B. Zhao and J. Tang, "Handwritten Digit Recognition Method Based on DTW and SVM," 2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qiangdao, China.

[10] X. Lv, "Handwritten Digit Recognition Based on Deep Learning Algorithms," 2023 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC), Rio De Janeiro, Brazil.