International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Identification of Handwritten Digit Using Deep Learning

Parrdhi Gupta[1], Mrs Pooja Chaudhary[2]

[1]Department of Computer Science and EngineeringAssistant Professor, Raj Kumar Goel Institute of TechnologyDepartment of Computer Science and EngineeringGhaziabad, India

[2]Raj Kumar Goel Institute of Technology, U.P. , India

Email: 707parrdhigupta@gmail.com , poojafds@rkgit.edu.in

Abstract

Promising outcomes have been observed in the fields of deep learning and machine learning algorithms. Automated systems are heavily relied upon for a variety of activities, ranging from object classification in photos to adding sound to silent films. Enabling computers to read and comprehend handwritten input from a variety of sources is known as handwritten text recognition, or handwriting recognition (HWR) or handwritten textrecognition (HTR). This survey study focuses on handwritten digit recognition using CNN, Multi-Layer Perceptron (MLP), and Support Vector Machines (SVM) models on the MNIST dataset. The main goal is to evaluate these models' accuracy and execution durations in order to determine which one, when taking computing efficiency into account, is best for properly identifying handwritten numerals.

Keywords: Convolutional Neural Network (CNN), Backpropagation, Epochs, Hidden Layers, Machine Learning, Deep Learning, Handwritten Digit Recognition, MNIST Dataset.

I. INTRODUCTION

Handwritten digit recognition permits computers to get it human-written numbers from different sources like pictures, documents, and touch screens. This process, which centers on recognizing digits from 0 to 9, has gathered broad research intrigued, particularly in the domain of deep learning. Practical applications incorporate permit plate acknowledgment, mail sorting, and preparing bank drafts. Present day strategies are essentially more progressed than early computer vision strategies, such as the neo-cognitron by Fukushima in 1980, which was affected by DH Hubel's work. Convolutional Neural Networks (CNNs) have demonstrated to be profoundly viable for picture investigation, contributing to areas such as question discovery, facial acknowledgment, video analysis, and design recognition. By utilizing angle plummet and backpropagation for preparing, CNNs have accomplished remarkable exactness in various assignments, counting manually written digit recognition, often reaching levels of accuracy comparable to human execution. When LeCun et al. 's seven-layer CNN classified handwritten digits straight from pixel images in 1998, it set a benchmark. The purpose of this study is to assess how different hidden layers in CNNs affect handwritten digit identification accuracy. We apply various CNN designs on the MNIST dataset using the TensorFlow framework, and then we evaluate their performance to identify the best model.

II. Literature Review

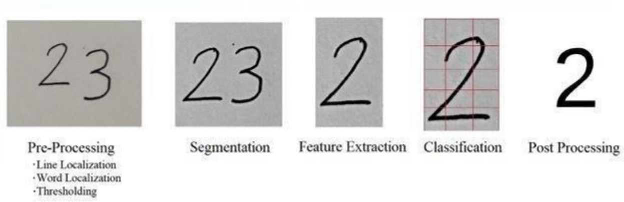
With the development of neural network topologies and early computer vision algorithms, the task of handwritten digit recognition has a long history. Because Support Vector Machines (SVM) are robust when processing high- dimensional data, they have been used for classification jobs for a long time. They function bylocating the best hyperplane in the feature spaceto divide several classes. A type of feedforward artificial neural network known as a multi-layer perceptron (MLP) is made up of several layers of nodes arranged in a directed graph, each of which is completely connected to the one above it. Backpropagation is used in MLP training to reduce classification mistakes. Although they work well, they frequently need a lot of hyperparameter tweaking and are not as good at processing image data as CNNs are. The capacity of Convolutional Neural Networks (CNNs) to naturally learn spatial progressions of highlights has revolutionized image processing. CNNs are particularly well-suited for applications like digit acknowledgment since they prepare input pictures utilizing convolutional layers, pooling layers, and completely associated layers. The 70,000 handwritten digit images in the MNIST dataset, a standard in this field, have been broadly utilized to evaluate different machine-learning procedures.

III. RELATED WORK

Convolutional Neural Networks (CNNs) are basic in numerous domains, especially image processing, and their employments are as it were developing. CNNs are crucial in nanotechnology, particularly in the semiconductor division, for recognizing and classifying generation imperfections. The acknowledgment of manually written numbers has earned expanding consideration in later times, as prove by the multiplication of research papers and publications on the subject. Inquire about appears that profound learning calculations outflank more customary machine learning methods like K-Nearest Neighbors (KNN), Random Forest Classifier (RFC), and Support Vector Machines (SVM) in terms of precision. This is particularly genuine for multi-layer CNNs that are executed utilizing frameworks like Keras with Theano and TensorFlow. Because of their excellent accuracy, CNNs are widely used in video analysis, image categorization, and other applications. Additionally, researchers are concentrating on sentiment recognition within words. They are experimenting with different parameters and using CNNs for natural language processing and emotional analysis. High performance in CNNs is getting harder to achieve since large neural networks require more parameters. Scientists are committed to reducing errors and enhancing CNN accuracy. Recent research has demonstrated that deep neural networks can perform better with simpler back-propagation training than with CIFAR-10, obtaining the lowest error rates on the MNIST dataset. For instance, using 3-NN trained and evaluated on MNIST, some researchers have achieved an excellent error rate of 1.19%. Deep CNNs are utilized for applications like sentence retrieval within images because they are very adaptive and reduce input image noise effectively. Researchers are working hard to create novel methods that get around the drawbacks of conventional convolutional layers. Using the MNIST dataset, Neural Convolutional Feature Map (NCFM), one such method, demonstrated applicability to large-scale data with an accuracy of 99.81% and notable performance gains. With an eye on lowering mistake rates, research on CNN applications is still being conducted. Researchers examine and examine error rates using datasets such as MNIST, and employ CNNs to rectify blurry images. As a consequence, new models with an accuracy of 98% and a loss range of 0.1% to 8.5% are proposed on the MNIST dataset.
A CNN-based model with 99.65% precision has been displayed for traffic sign distinguishing proof in Germany, illustrating speedier execution. Misfortune capacities that are particularly

IV. METHODOLOGY

Pre-processing, segmentation, include extraction, classification, and recognition are a few of the fundamental steps that make up handwriting recognition. Post-processing comes final.

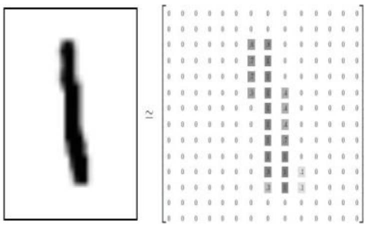


A few processes, including thresholding, word localization, and line localization, are carried out amid the preprocessing stage. These methods set up the information for extra examination. Words are ordinarily part into portions that compare to particular letters or numbers amid the segmentation phase. Be that as it may, since the MNIST dataset as of now includes pre-segmented images of numbers, the segmentation phase was excluded from this examination.The feature extraction phase includes changing the data into a more compact and enlightening representation, encouraging the consequent recognition step. The classification and acknowledgment stage ordinarily includes allotting typical classes to the characters. Different methods can be utilized in this phase, including artificial neural networks (ANN) and other classification algorithms. The choice of dataset is vital for this phase, as the differing qualities and abundance of the preparing data altogether affect the algorithm's success rate. In the post-processing stage, the objective is to rectify any potential mistakes that may have happened amid the recognition handle. A few frameworks execute dictionaries at this organize, where the recognized content is confirmed against a list of conceivable words. In this think about, the MNIST dataset was utilized, which is as of now well-suited for written by hand digit recognition errands, meaning that comes about might change altogether if a diverse dataset were utilized. Written by hand digit recognition is influenced by different components andconditions.

Data Preparation

Our studies are based on the MNIST dataset, which comprises of 10,000 testing images and 60,000 training images of handwritten digits. The digits in each picture are centered and size-normalized within a 28x28 grayscale pixel framework. Normalizing pixel values to the range [0, 1] and molding the information to meet the input determinations of different models are examples of preprocessingprocedures.

Sample images of MNIST data



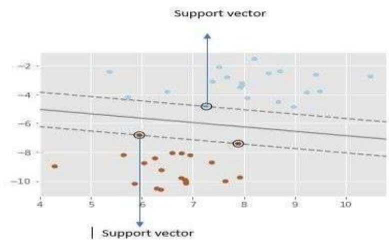
The digit 1 in MNIST can be represented in matrix form

- Precision: TP / (TP + FP). It represents the accuracy rate of positive predictions.
- Recall: TP / (TP + FN) It indicates the sensitivity or true positive rate.
- F1 Score: This value is the harmonic mean of precisions.
- Accuracy: Also known as accuracy rate, it reflects the accuracy of the predictions.

Model Implementation

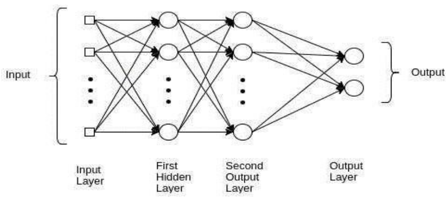
A. Support Vector Machines (SVM)

To categorize the digit pictures, we utilize a linear kernel to implement the SVM. The MNIST training set is utilized to train the model, whereas the test set is utilized to survey it. Grid search and cross-validation areutilized to optimize hyperparameters like the regularization parameter.



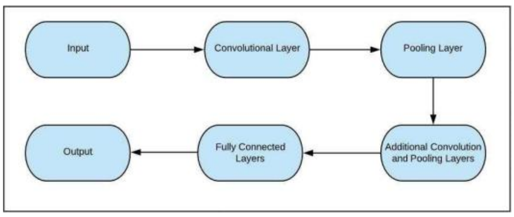
B. Multi-Layer Perceptron (MLP)

The MLP model comprises of an input layer, two hidden layers with ReLU activation functions, and an yield layer with a softmax work. We utilize stochastic gradient descent (SGD) for training, altering learning rates and batch sizes to optimize execution. The model is trained for a settled number of epochs, and early ceasing is utilized to prevent overfitting.



C. Convolutional Neural Network (CNN)

Our CNN architecture incorporates convolutional layers with ReLU actuations, max-pooling layers, and completely associated layers. The network is trained utilizing the Adam optimizer, and hyperparameters such as channel sizes, number of channels, and layer profundities are fine-tuned through experimentation.Dropout is applied to moderate overfitting, and the model is approved utilizing the test set.



Evaluation Metrics

We utilize accuracy as the essential metric to assess model execution. Execution time for training and deduction is moreover recorded to survey computational proficiency. Confusion matrices are created to visualize demonstrate predictions and identify common misclassifications.

V. Results

Support Vector Machines (SVM)

The SVM model accomplished an exactness of 92.5% on the MNIST test set. Whereas the classification performance is strong, the training time was generallylong due to the high dimensionality of the input information.

Multi-Layer Perceptron (MLP)

The MLP model yielded an precision of 97.1%, outflanking the SVM in terms of classification precision. Training time was altogether diminished compared to SVM, highlighting the effectiveness of neural network-based approaches.

Convolutional Neural Network (CNN)

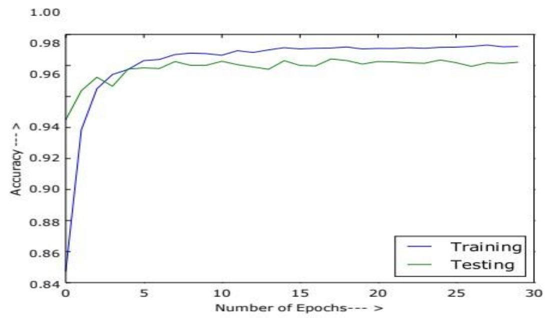
The CNN model accomplished the most elevated accuracy of 99.2%, illustrating its predominance indealing with image information. The training time was comparable to the MLP, but the induction time was marginally longer due to the complexity of the CNN architecture. The perplexity matrix demonstrated negligible misclassifications, essentially between outwardly comparable digits.

The overall performance of each model is shown in this table. The table has five columns: the name of the model appears in the second, training and testing accuracy are shown in the third and fourth columns, and model execution time isshown in the fifth column.

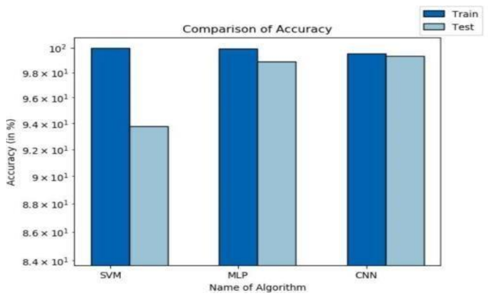
TABLE

COMPARISON ANALYSIS OF DIFFERENT

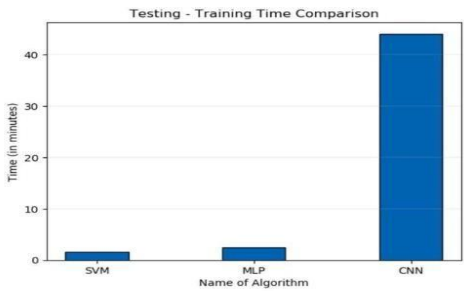
SAMPLE TRAINING % TESTING % IMPLEMENTATION



(Accuracy v/s Number of epochs).

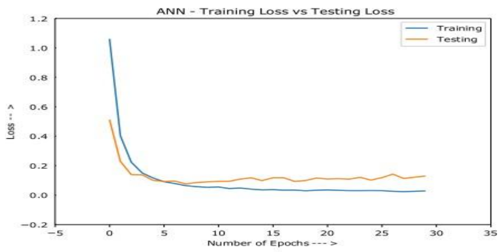


Bar chart demonstrating accuracy comparison



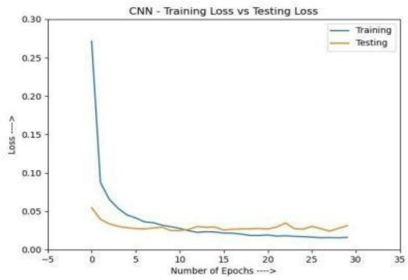
A bar chart comparing the SVM, MLP, and CNN execution times is displayed.

Additionally, we showed how deep learning modelsimproved accuracy and decreased error ratesconcerning the number of epochs through performance metrics. Drawing the chart is important because it helps us determine whether to apply an early stop, which preventsoverfitting, which occurs when the accuracy changes steadily over time.



(rate of loss V/S epoch count).

SVM	99.89%	94.05%	1:36 MIN
MLP	99.72%	98.75%	2:33 MIN
CNN	99.63%	99.21%	44:07 MIN



Graph showing the CNN training loss transition withincreasing number of epoch (Loss rate v/s Number of epochs).

VI. Discussion

The results demonstrate that CNNs are the most effective model for handwritten digit acknowledgment, accomplishing the most elevated exactness and illustrating robustness in image classification tasks. Whereas SVMs and MLPs offer competitive execution, CNNs' capacity to automatically extricate spatial highlights from pictures gives a particular advantage. The longer training times for SVMs can be credited to therequire to optimize the hyperplane in high-dimensional space, whereas MLPs and CNNs advantage from parallelization in modern deep-learning systems. The choice of demonstrate ought to consider the particular application prerequisites, balancing accuracy and computational resources.

VII. Conclusion

This study highlights the viability of CNNs for handwritten digit acknowledgment utilizing the MNIST dataset, outperforming SVM and MLP models in terms of precision and computationalproficiency.

Future work can investigate progressed architectures, such as more profound networks and exchange learning, to assist improve execution. The discoveries emphasize the significance of selecting fitting models for particular assignments, leveraging the strengths of deep learning for practical applications.

Future Enhancement

Future enhancements in the field of handwritten digit recognition seem center on a few key regions to progress exactness and proficiency. One potential course is the investigation of more progressed profound learning structures, such as more profound Convolutional Neural Systems (CNNs) or novel neural arrange plans, which might give superior highlight extraction and classification capabilities. Incorporating techniques like exchange learning, where pre-trained models on expansive datasets are fine-tuned for particular errands, seem moreover upgrade execution. Also, leveraging outfit strategies, which combine different models to improve predictive accuracy, might be useful. Another area for improvement is the utilize of more assorted and broad datasets for training. This seem include datasets that incorporate varieties in penmanship styles, diverse dialects, and more complex digit arrangements. Expanding the differences of training information can offer assistance the models generalize superior to real- world scenarios. Improving preprocessing and information augmentation methods can moreover lead to way better acknowledgment rates. Strategies such as picture normalization, clamor lessening, and the presentation of engineered information through expansion can offer assistance make more vigorous models. Furthermore, integrating contextual data and natural language processing (NLP) methods in the post-processing stage can diminish blunders. For instance, utilizing dictionaries or dialect models to approve recognized digits inside a word or sentence setting can make strides exactness. Lastly, upgrading the computational efficiency of these models through optimization strategies and the utilize of specialized hardware, like GPUs and TPUs, can make the arrangement of these models more commonsense in real-time applications. By centering on these ranges, future investigate can proceed to thrust the boundaries of what is conceivable in handwritten digit recognition, leading to more exact and solid systems.

REFERENCES

- Alpaydin, E., & Kaynak, C. (1998). Optical recognition of handwritten digits: Data set. University of California, Irvine, School of Information and Computer Sciences.
- Zhang, Y., & LeCun, Y. (2015). Text understandingfrom scratch. arXiv preprint arXiv:1502.01710.
- Deng, L., & Yu, D. (2014). Deep learning: methodsand applications. Foundations and Trends® in Signal Processing, 7(3–4), 197-387.
- Wan, L., Zeiler, M., Zhang, S., LeCun, Y., & Fergus,
- R. (2013). Regularization of neural networks using dropconnect. In International Conference on MachineLearning (pp. 1058-1066). PMLR.
- Tang, Y. (2013). Deep learning using linear supportvector machines. arXiv preprint arXiv:1306.0239.
- Hochreiter, S., &Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
- Ioffe, S., & Szegedy, C. (2015). Batch normalization:Accelerating deep network training by reducing internalcovariate shift. In International conference on machine learning (pp. 448-456). PMLR.
- Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. In Proceedings of the twenty-first international conference on Machine learning (p. 78).
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXivpreprint arXiv:1406.1078.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Johnson, R., & Zhang, T. (2016). Supervised and semi-supervised text categorization using LSTM forregion embeddings. In International Conference on Machine Learning (pp. 526-534). PMLR.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... &Polosukhin, I. (2017). Attention is all you need. Advances in neural informationprocessing systems, 30.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P.(1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
- Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In Icdar (Vol. 3, pp. 958-962).
- Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine, 29(6), 141-142.
- Hinton, G. E., &Salakhutdinov, R. R. (2006).Reducing the dimensionality of data with neuralnetworks. Science, 313(5786), 504-507.
- Ciresan, D. C., Meier, U., Masci, J., Gambardella,
- L. M., &Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. In IJCAI Proceedings-International Joint Conference on Artificial Intelligence (Vol. 22, No. 1, p. 1237).
- Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. In Proceedings of the 30th InternationalConference on Machine Learning (ICML-13) (pp.1319-1327).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.
- Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In Internationalconference on artificial neural networks (pp. 92- 101). Springer, Berlin, Heidelberg.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. r., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modelingin speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine,29(6), 82-97.
- Fukushima, K. (1980). Neocognitron: A self- organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36, 193-202.
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., &Salakhutdinov, R. (2014). Dropout:a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.