

Handwritten Character Recognition Using CNN

¹Dr. V. Suganthi, ²S. Sankeerthna

¹Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamilnadu, India

²Student, Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamilnadu, India

Abstract - Handwritten character recognition holds a significant position in the realms of pattern recognition and machine learning, finding applications across archival document digitization and automated form processing. Convolutional Neural Networks (CNNs) have emerged as a powerful tool for addressing this task due to their ability to learn intricate features and patterns from raw input data. This paper provides a comprehensive overview of CNNs' application in handwritten character recognition, covering various aspects of the recognition process including preprocessing techniques, dataset selection, network architecture design, training methods, and performance evaluation. The synthesis explores the evolution of CNN-based methodologies, examining architectural enhancements, regularization techniques, and data augmentation methods that collectively improve recognition accuracy. Additionally, the paper evaluates advanced approaches that integrate CNNs with other machine learning techniques such as recurrent networks and attention mechanisms, leading to further improvements in recognition performance. Through a thorough review of contemporary literature, the paper highlights achievements, ongoing challenges, and emerging research directions in the field of handwritten character recognition using CNNs. This survey serves as a valuable resource for researchers and practitioners interested in understanding the current landscape of the field and sheds light on potential breakthroughs and the trajectory of advancements in this dynamic domain.

Keywords: Handwritten character recognition leveraging Convolutional Neural Networks: advancements, challenges, and emerging trends.

I. INTRODUCTION

In an increasingly digitized world, the ability to automatically recognize and transcribe characters from various sources has become a fundamental technology with widespread applications. Whether in the context of Optical Character Recognition (OCR) for digitizing printed documents, automated data entry in forms, or multilingual text processing, character recognition plays a pivotal role in enhancing efficiency, accuracy, and accessibility. The task of accurately identifying and classifying characters, particularly

in handwritten or stylized forms, presents intricate challenges that have driven the development of advanced machine learning techniques. Convolutional Neural Networks (CNNs) have emerged as a cornerstone in the field of computer vision and pattern recognition, offering unprecedented capabilities in image analysis and classification. Character recognition, a subdomain of computer vision, has witnessed substantial advancements thanks to CNNs, which excel at capturing intricate patterns and features from pixel-level data. By leveraging hierarchical layers of convolution and pooling operations, CNNs have demonstrated remarkable success in handling the nuances and complexities of character recognition tasks. In this context, the art of expression takes many forms, from the written word to visual and mediums[3]. It encompasses language, both spoken and written, as well as the visual arts, music, and the myriad ways we share our thoughts, emotions, and experiences. Communication bridges gaps, illuminates the unknown, and fosters connections among individuals and communities, transcending geographical, cultural, and linguistic boundaries.

II. REVIEW OF LITERATURE AND RELATED WORKS

This section offers an overview of the significant strides made in the field of handwritten character and digit recognition through the lens of recent academic contributions. The evolution of machine learning techniques, from the use of shallow networks to more complex models like deep belief networks (DBNs) and recurrent neural networks (RNNs), has led to notable advancements in accuracy and reliability [1]. Specifically, Hinton and colleagues demonstrated the effectiveness of DBNs by achieving an accuracy rate of nearly 98.75% on the widely used MNIST dataset. Furthermore, Pham and their team enhanced the capabilities of RNNs by incorporating dropout regularization, thus paving the way for more robust recognition of unconstrained handwriting. The advent of convolutional neural networks (CNNs) marked a turning point in handwritten character recognition (HCR), with researchers achieving unprecedented accuracy levels.[4] The groundwork for this success was laid by Simard and others, who in 2003 introduced a seminal CNN architecture tailored for visual document analysis. Subsequent studies, such as those by Wang et al., leveraged multilayer CNNs for

comprehensive text recognition tasks, showcasing the model's effectiveness across various benchmark datasets.[7]

A novel approach that combined the strengths of deep CNNs (DCNN) and RNNs was introduced by Shi et al. through the development of the Conventional Recurrent Neural Network (CRNN). This method outperformed traditional recognition techniques, especially in scene text recognition. Similarly, the introduction of SegNet by Badrinarayanan et al. represented a significant leap forward in semantic segmentation, offering superior performance thanks to its deep convolutional network framework [2].

The versatility of CNNs was further evidenced by their successful application in offline handwritten character recognition across different languages. For instance, Gupta and colleagues achieved an impressive accuracy rate of 95.96% in recognizing a mix of Bangla numerals, Devanagari characters, and English numerals, underscoring the model's adaptability. In another study focusing on the CROHME dataset, Nguyen et al. employed multiscale CNNs to achieve high recognition performance.

The challenge of identifying word locations in historical texts was addressed by Ziran et al. through an R-CNN-based framework, demonstrating the model's utility in handling complex, real-world documents. Meanwhile, Ptucha et al. developed an intelligent character recognition (ICR) system utilizing conventional neural networks, yielding noteworthy results.[8]

The critical role of hyper-parameter optimization in maximizing the efficiency and accuracy of CNNs was highlighted, with the understanding that inappropriate parameter selection can significantly hinder model performance. The exploration of different CNN models revealed a variance in the number of hyper-parameters, emphasizing the importance of meticulous adjustment by researchers to enhance recognition capabilities.[3]

Lastly, the work of Tapotosh Ghosh and colleagues involved experimenting with advanced image processing techniques using models such as InceptionResNetV2, DenseNet121, and InceptionNetV3 on the CMATERdb dataset. The forthcoming section, Results and Analysis, promises a detailed examination of these research endeavors, including dataset preprocessing and methodological approaches to optimize recognition accuracy

III. DATASET

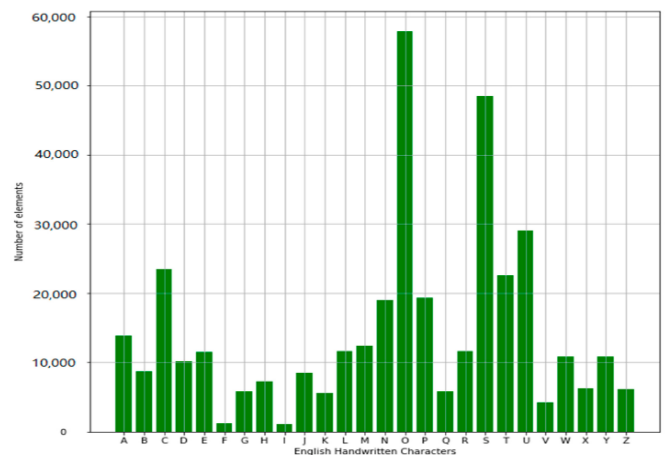
The MNIST digit benchmark dataset, derived from the National Institute of Standards and Technology (NIST), consists of two categories: digits and alphabets. It is easily

accessible through Keras functionality, comprising 60,000 training examples and 10,000 test examples [59]. However, for this project's digit recognition task, only the dataset containing 10 classes of MNIST digits is used.

Each digit in the dataset is standardized to a uniform size, and its binary image is centered into a 28×28 format by computing the center of mass of the pixels. The test set comprises 5000 patterns sourced from two datasets, each containing approximately 30,000 patterns contributed by about 250 different writers, including high school students and Census Bureau employees [1]. The datasets are labeled accordingly for easier verification.

The Kaggle alphabet dataset is compiled from various sources, including NIST, NMIST, and Google Images [60]. It features English handwritten alphabets categorized into 26 classes, with over 297,000 training examples and a test set of over 74,490 examples. The distribution of Kaggle letters is depicted in Figure 4.

Similar to the MNIST dataset, each letter's binary image in the Kaggle alphabet dataset is standardized to a uniform size, with its center of mass computed to ensure it is centered into a 28×28 format.



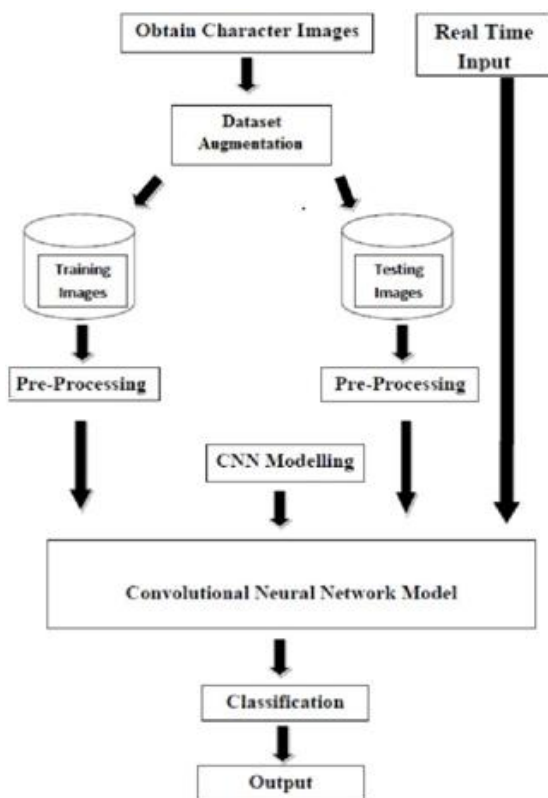
IV. PROPOSED CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNNs) have gained immense popularity for their exceptional performance in image classification tasks, offering numerous applications in various fields like medical imaging, traffic monitoring, and facial recognition. CNNs, inspired by the functioning of the visual cortex, are composed [4] of artificial neurons with self-optimization capabilities, enabling precise feature extraction and classification from images with minimal preprocessing[5].

In CNNs, each pixel in an image serves as a feature for the neural network, where low-level features like edges and

gradients are captured in the initial convolutional layers. As the network deepens, it extracts higher-level features, leading to increased computational complexity. However, advancements in computational technology and algorithmic development have mitigated this issue, making CNNs more efficient.

Handwritten character recognition (HCR) with CNNs has been an early focus of research in deep learning. Despite significant improvements in modeling efficacy and access to extensive datasets, ensuring model generalization across diverse datasets remains a challenge. To address this, a computationally efficient CNN model is proposed, optimized with a limited number of convolutional layers while maintaining high accuracy across multiple datasets.[11]



V. RESULTS AND ANALYSIS

In our handwritten recognition endeavor, we engaged two datasets: Kaggle for English letters (A–Z) and MNIST for numeric characters (0–9). Employing both 'ADAM' and 'RMSprop' optimizers, we traversed a spectrum of learning rates—0.001, 0.0001, and 0.00001—resulting in the creation of six CNN models for each dataset, totaling twelve models[10].

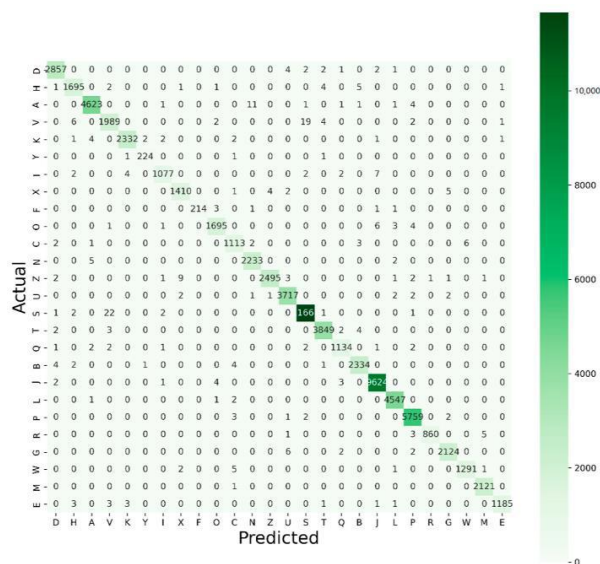
To streamline communication and avoid redundancy, we assigned distinct names to each model. For Kaggle, models with LR 0.001 were dubbed K1 ('ADAM') and K2 ('RMSprop'). For LR 0.0001, they were K3 ('ADAM') and K4

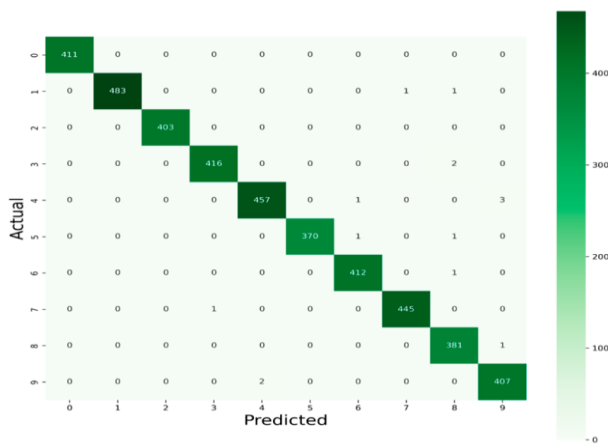
('RMSprop'). Lastly, LR 0.00001 yielded K5 ('ADAM') and K6 ('RMSprop'). A similar nomenclature scheme was adopted for the MNIST dataset, labeled from M1 to M6. After rigorous evaluation; we uncovered optimal performance under 'ADAM' with LR 0.00001 for Kaggle (K5) and 'RMSprop' with LR 0.001 for MNIST (M2). Following this discovery, we proceeded to compute F1 scores (micro, macro, and weighted average), craft confusion matrices, and compile two classification reports for these top-performing models on each dataset.

When assessing model performance, clarity emerges from matrices detailing specificity, recall, precision, F1 score, and support. These metrics, derived from the confusion matrix (CM), provide insight into the model's capabilities. The CM categorizes outcomes into four distinct results: total false positive (TFP), total false negative (TFN), total true positive (TTP), and total true negative (TTN). With these values, it becomes feasible to compute per-class metrics such as recall, precision, specificity, and F1 score for each dataset.[9]

Consider a scenario where our goal is to detect the letter 'A'. For simplicity, let's assume there are 100 images for each of the 26 letters in the alphabet, totaling 2600 images. If the model correctly identifies 'A' in 97 out of 100 images, the accuracy of the model is 97%, indicating 97 true positives (TTPs). If, under the same assumptions, the letter 'O' is mistakenly identified as 'A', we would observe 1 false positive (TFP). Conversely, if 'A' is misclassified as 'O' three times, the model incurs 3 false negatives (TFNs). The remaining 2499 images out of the total 2600 are deemed true negatives (TTN)[8].

These models were selected based on overall performance, trained and validated using the Kaggle and MNIST datasets, respectively.





VI. CONCLUSIONS

In contemporary times, handwritten character recognition (HCR) systems are witnessing a surge in applications. In this study, we propose a CNN-based model tailored for multiclass classification in HCR systems, achieving remarkable results. Our models were trained using the MNIST digit dataset, comprising 60,000 training and 10,000 testing images, and the extensive Kaggle alphabet dataset, boasting over 297,000 training images and a test set of 74,490 images.

For the Kaggle dataset, employing the 'ADAM' optimizer yielded overall accuracies of 99.516%, 99.511%, and 99.563% with learning rates (LR) of 0.001, 0.0001, and 0.00001, respectively. Meanwhile, using 'RMSprop', the accuracies were 99.292%, 99.108%, and 99.191% for the corresponding LR. Conversely, for the MNIST dataset, 'RMSprop' achieved accuracies of 99.642%, 99.452%, and 98.142%, while 'ADAM' yielded 99.571%, 99.309%, and 98.142% accuracies for the same LR.

Notably, for alphabet recognition, accuracy declined as LR increased, whereas in digit recognition, accuracy exhibited a proportional relationship with LR. Precision, recall, specificity, and F1 score were computed from confusion matrices. Among the twelve models discussed, the 'ADAM' optimizer with LR 0.00001 achieved a recall value of 99.56%, while the 'RMSprop' optimizer with LR 0.001 attained a recall value of 99.64%, distinguishing these two models for the Kaggle and MNIST datasets, respectively[6].

Given the imbalanced dataset distribution, relying solely on accuracy for model evaluation proves ineffective. Therefore, classification reports (CR) were included, showcasing F1 scores for every 10 classes for digits (0-9) and every 26 classes for the alphabet (A-Z).

Furthermore, we compared the results of our best two models with other notable works in the field. Looking ahead, future work will explore integrating various feature extraction

methods within a similar framework to address more complex languages like Korean, Chinese, Finnish, and Japanese.

REFERENCES

- [1] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495.
- [2] Gupta, P., Majumdar, A., & Gupta, D. (2019). CNN based approach for multilingual handwritten character recognition. In 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1460-1464). IEEE.
- [3] Ghosh, T., & Amin, M.A. (2020). Handwritten Bangla character recognition using deep learning with Transfer Learning. In 2020 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-4). IEEE.
- [4] Hinton, G.E., Osindero, S., & Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554.
- [5] Nguyen, C.H., Dang, K., & Doermann, D. (2019). Multiscale CNNs for offline handwritten mathematical expression recognition. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 1, pp. 7441-7448).
- [6] Ptucha, R., & Savakis, A.E. (2013). A new approach to intelligent character recognition using computational intelligence. In 2013 IEEE International Conference on Systems, Man, and Cybernetics (pp. 3961-3966). IEEE.
- [7] Pham, V.T., Bluche, T., Kermorvant, C., & Louradour, J. (2014). Dropout improves recurrent neural networks for handwriting recognition. In *Frontiers in Handwriting Recognition* (pp. 285-290). IEEE.
- [8] Shi, B., Bai, X., & Yao, C. (2016). 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*, 39(11), 2298-2304.
- [9] 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). IEEE.
- [10] Wang, K., Babenko, B., & Belongie, S. (2011). End-to-end scene text recognition. In 2011 International Conference on Computer Vision (pp. 1457-1464). IEEE.
- [11] Ziran, Z., Lin, H., Huang, S., & Xiong, H. (2020). Word location recognition in historical document images using R-CNN-based framework. *Journal of Imaging Science and Technology*, 64(5), 050501.

Citation of this Article:

Dr. V. Suganthi, S. Sankeerthna, "Handwritten Character Recognition Using CNN", Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 3, pp 194-198, March 2024. Article DOI <https://doi.org/10.47001/IRJIET/2024.803026>
