Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

A Deep Learning Approach to Recognize Bengali Handwritten Digit Using Transfer Learning

Md. Ryhan Uddin[1], Md. Sadikur Rahman Khan[2], Tanzeem Tahmeed Reza[3], Sakib Ahmed[4], Mohammad Sohel[5], and Md. Abdullah[6]
Department of Electronics and Telecommunication Engineering,
Chittagong University of Engineering and Technology, Chattogram-4349, Bangladesh
([1]ryhan.cuet, [2]sadik.cuet, [3]tanzeemtahmeed01, [4]ahmedalvi5418, [5]sohelete18, [6]abdullahmymon0)@gmail.com

***Abstract*—This paper presents a robust and efficient approach for Bengali handwritten digit recognition using transfer learning model. The original RGB dataset is converted to grayscale to reduce complexity, followed by preprocessing to replicate the grayscale channel three times to align with the RGB input requirements of pretrained models. Leveraging transfer learning, the InceptionResNet model is fully retrained with custom dense and dropout layers to optimize feature extraction and classification. Extensive experimentation on a large dataset yielded exceptional performance metrics, achieving 98.87% training accuracy, 98.72% validation accuracy, and 98.26% testing accuracy. The proposed approach significantly outperforms traditional methods and contemporary deep learning architectures, offering a reliable solution for automated Bengali digit recognition. This study underscores the effectiveness of transfer learning and fine-tuning in addressing challenges posed by complex handwritten scripts, establishing a foundation for future advancements in multilingual handwritten character recognition.**

***Index Terms*—Bengali Handwritten Digits, InceptionResNet, Deep Learning, Image Classification, Pattern Recognition.**

I. Introduction

Handwritten digit recognition is a critical component of optical character recognition (OCR) systems, with widespread applications in automated banking, postal sorting, academic grading, and digitized record management. While significant advancements have been made in recognizing numerals from widely used scripts like Latin and Arabic, regional scripts such as Bengali pose unique challenges. The Bengali script, widely used in Bangladesh and parts of India, is characterized by its complex shapes, curved strokes, and contextual intricacies, making accurate recognition a demanding task. Given that over 230 million people rely on Bengali script for communication, developing robust digit recognition systems is essential for bridging socio-economic gaps and fostering technological inclusivity.

Traditional methods for handwritten digit recognition have focused on handcrafted features combined with machine learning algorithms. While these methods achieved notable accuracy on structured datasets, they often struggled with the variability inherent in real-world handwritten samples, such as inconsistent writing styles, noise, and distortions. Additionally, these approaches required significant domain expertise for feature engineering, limiting their adaptability to diverse datasets and scripts.

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have transformed image recognition tasks. CNNs automatically extract hierarchical features, enabling robust recognition even in challenging scenarios. Transfer learning, a popular deep learning paradigm, leverages pretrained models on large datasets and fine-tunes them for specific tasks, reducing computational requirements and improving performance on smaller datasets.

In this study, we present a fine-tuned InceptionResNet model for Bengali handwritten digit recognition [1]. By converting the RGB dataset to grayscale, we reduce computational complexity while retaining critical image features. The grayscale images are then adapted for pretrained models, which typically require three-channel inputs, by replicating the grayscale channel. Our model combines transfer learning with customized layers to achieve high accuracy and generalization.

The remainder of this paper is organised as follows: Section II discusses related works. Our methodology is introduced in Section III. Section IV details our experiments and findings. Finally, Sections V and VI present our discussion and future research directions.

II. Related Works

Bengali handwritten digit recognition has been extensively studied in the literature, with numerous approaches proposed to address the challenges associated with diverse handwriting styles and character variations [2]. Traditional methods often rely on feature engineering and classical machine learning algorithms, while more recent approaches leverage deep learning techniques for improved performance and generalization.

One common approach in traditional methods is to extract handcrafted features from digit images, such as histogram of oriented gradients (HOG) or local binary patterns (LBP), followed by classification using classifiers like SVM or k-NN [3], [4]. These methods have been widely used and have shown reasonable performance, especially in constrained environments with limited variability in handwriting styles. However, they often require domain expertise in feature engineering and may struggle to generalize well to diverse datasets with complex variations.

In recent years, deep learning techniques have emerged as state-of-the-art methods for handwritten digit recognition. CNNs have shown remarkable success in image classification tasks, including digit recognition [5]. More advanced CNN architectures have further improved the state-of-the-art performance in image classification tasks [6]. These architectures typically consist of multiple convolutional layers followed by pooling layers and fully connected layers for feature extraction and classification. However, these models may suffer from overfitting, especially when trained on limited datasets.

To address overfitting and improve generalization, transfer learning has become increasingly popular in recent years. Transfer learning involves leveraging pretrained models trained on large-scale datasets and fine-tuning them on target datasets for specific tasks [7]. By transferring knowledge learned from one domain to another, transfer learning enables effective feature extraction and improves model performance, especially when training data is limited.

III. Methodology

In this section, we outline the methodology employed in the development and training of our deep learning model for recognition of handwritten digits. The methodology encompasses data preprocessing, model architecture selection, and training strategy.

A. *Dataset*

The dataset utilized in this study is sourced from the NumtaDB dataset, which contains a substantial collection of 180x180 RGB images of Bengali handwritten digits [8]. The dataset is divided into three subsets: 65% for training, 15% for validation, and 20% for testing. This split ensures that the model is trained on a diverse set of examples and evaluated on a sufficiently large validation set to gauge its performance accurately. The test set is reserved for final evaluation to assess the model's generalization to unseen data.

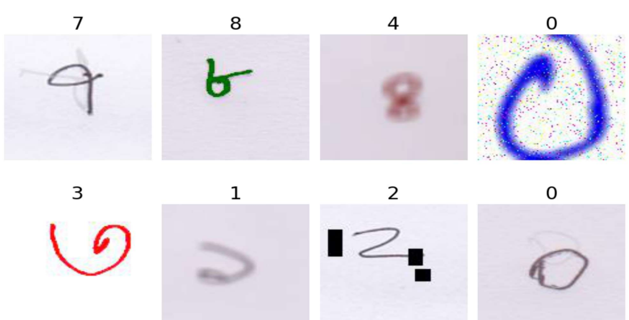

Fig. 1: Example images from the NumtaDB dataset.

B. *Dataset Preprocessing*

To ensure optimal model performance, we preprocessed the dataset by resizing images to 75x75 pixels and converting them to grayscale. Grayscale images were replicated along the channel axis to create 3-channel tensors required by InceptionResNet. Pixel values were normalized to the [0, 1] range by dividing by 255, ensuring faster convergence. These steps standardized the input dimensions and preserved essential features for effective digit recognition.

C. *Model Architecture*

For the task of digit recognition, we employed a transfer learning approach by retraining the InceptionResNet architecture. InceptionResNet is a deep convolutional neural network pre-trained on the ImageNet dataset and is well-known for its outstanding performance on image classification tasks. Leveraging transfer learning, we fine-tuned the model to accommodate the specific characteristics of handwritten digits.

The final architecture (Figure 2) of the model consisted of the following components:

- **Base Model:** InceptionResNet retrained on the training set, serving as the feature extractor. This allowed us to capture high-level features from the input images effectively.
- **Global Average Pooling Layer:** To reduce the dimensionality of the extracted features and obtain a fixed-length feature vector for each image, a global average pooling layer was employed.
- **Fully Connected Layers:** After the global average pooling layer, multiple fully connected layers with ReLU activation functions were added to capture intricate patterns and facilitate non-linear transformations. A bottleneck layer with 64 neurons compresses essential features into a lower-dimensional representation, which is then expanded through additional layers. To prevent overfitting, dropout layers were incorporated between the fully connected layers, with dropout rates of 0.20, 0.20, and 0.10 for the respective layers.
- **Output Layer:** A new output dense layer with 10 neurons, followed by a softmax activation function, enables the classification of Bangla handwritten digits, aligning with the objectives of this study.

D. *Model Optimization*

The model was optimized with the hyperparameters detailed in Table I. Early stopping was applied to prevent overfitting, ensuring efficient convergence and optimal performance in digit recognition.

TABLE I: Hyperparameters in Proposed Model Architectures

Hyperparameter	Value/Type
Input Size	75 x 75 x 3
Epochs	10
Batch Size	64
Optimizer	Adam
Learning Rate	1×10^{-3}
Loss Function	Categorical Cross-entropy

IV. Experiments and Findings

This section presents the experimental setup, results, and analysis for Bengali handwritten digit recognition, including a confusion matrix, classification report, and comparison with existing methods.

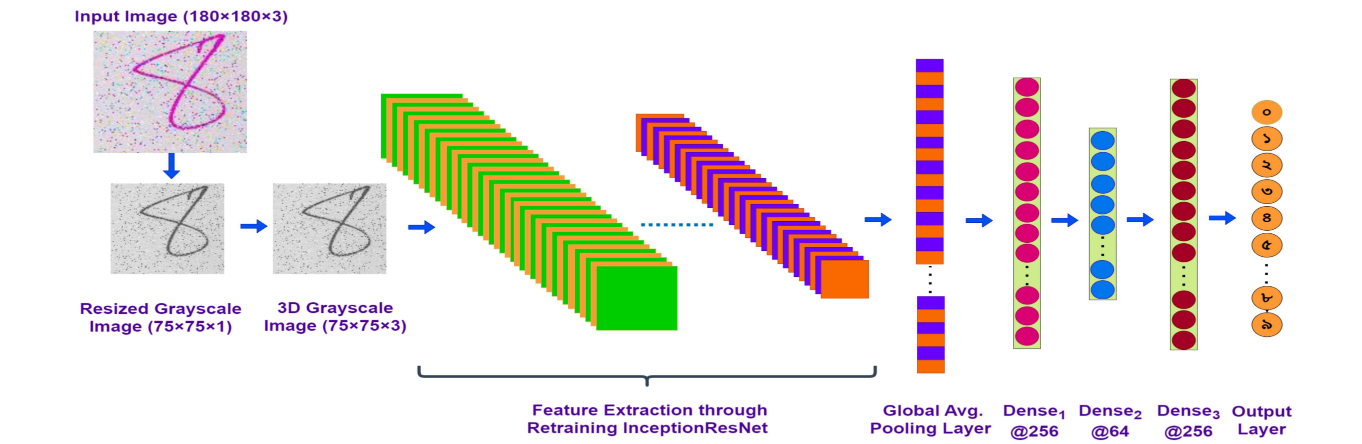

Fig. 2: A comprehensive schematic depicting the layered structure and intricate design of the retrained InceptionResNet architecture, showcasing key components optimized for accurate prediction in Bengali handwritten digit recognition.

A. *Experimental Setup*

The experiments were conducted on the Google Colab platform, leveraging a 13 GB RAM, 222 GB Colab Disk, and an Nvidia® Tesla T4 GPU with CUDA V11.2 for GPU acceleration. The setup utilized Keras® 2.4.3 API with TensorFlow® 2.5.0 backend and Python® 3.7.10 for model implementation. Visualization was performed using the matplotlib® 3.2.2 library.

B. *Results and Analysis*

The confusion matrix (Figure 3) highlights the model's high accuracy in classifying Bengali digits, with most digits correctly identified. However, some misclassifications, particularly among visually similar digits, slightly impacted the precision and recall for digit '1'. Despite this, the model demonstrates strong overall performance.

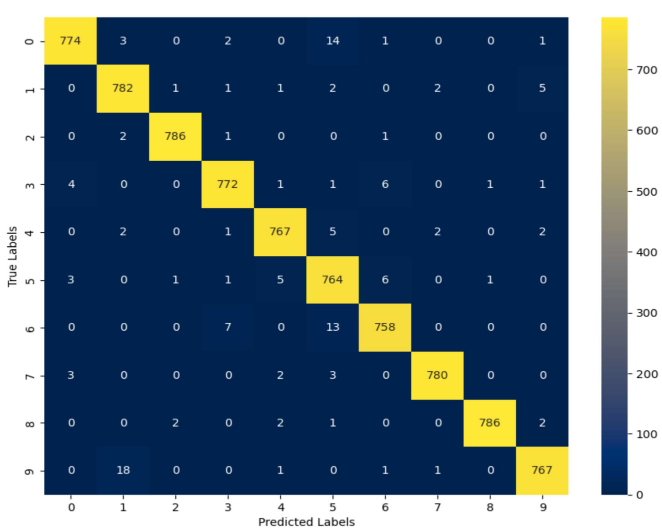

Fig. 3: The confusion matrix shows the model's accuracy, with correct predictions on the diagonal and misclassifications off-diagonal.

The classification report (Figure 4) summarizes the model's performance metrics for Bengali handwritten digit recognition. Precision scores, ranging from 0.95 to 0.99, reflect the proportion of correctly predicted instances for each class, while recall values (0.97 to 0.99) indicate the model's ability to identify most instances from each class. F1-scores, balancing precision and recall, range from 0.96 to 0.99, demonstrating consistent performance across classes. The model achieved an impressive accuracy of 98.26%, showcasing its robustness and reliability. These results highlight the effectiveness of deep learning techniques, particularly fine-tuning pretrained models like InceptionResNet, for accurate and generalizable image classification tasks.

Figure 5 displays Bengali handwritten digit predictions, showing high confidence for correct predictions and low confidence for incorrect ones, offering insight into the model's accuracy and its struggle with more difficult cases.

C. *Result Comparison*

We compare our method with existing approaches for Bengali handwritten digit recognition. Achieving 98.26% accuracy on the NumtaDB dataset, our method outperforms traditional techniques and deep learning models like CNN, Ensemble CNN, and VGG16, highlighting its superiority in Bengali digit.

TABLE II: Comparison of Results between the Proposed method and Existing approaches.

Model	Dataset	Accuracy
SVM + HOG [3]	CMATERdb 3.1.1	98.05
LBP + KNN [4]	CMATERdb 3.1.1	96.7
CNN [5]	NumtaDB	92.72
Ensemble CNN [6]	NumtaDB	96.79
VGG16 [7]	NumtaDB	97.09
Proposed	**NumtaDB**	**98.26**

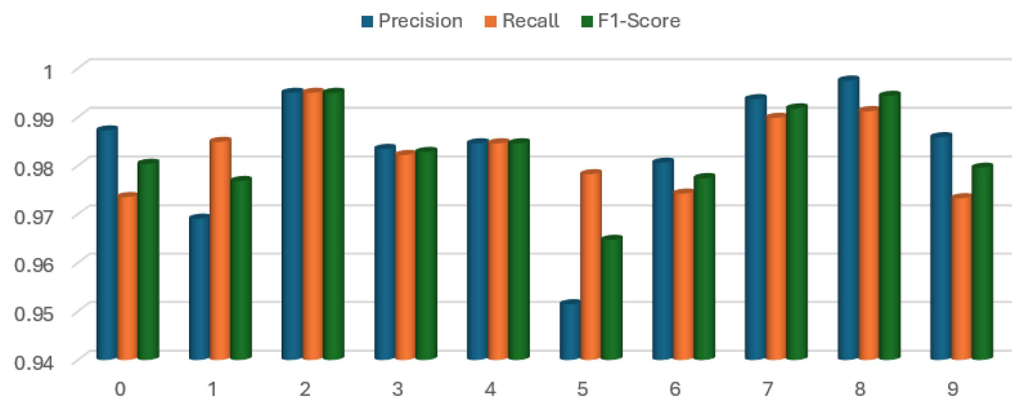

Fig. 4: The detailed classification report for the proposed model shows precision, recall, F1-score, and accuracy for each class. The high accuracy achieved underscores the model's overall effectiveness on the testing data.

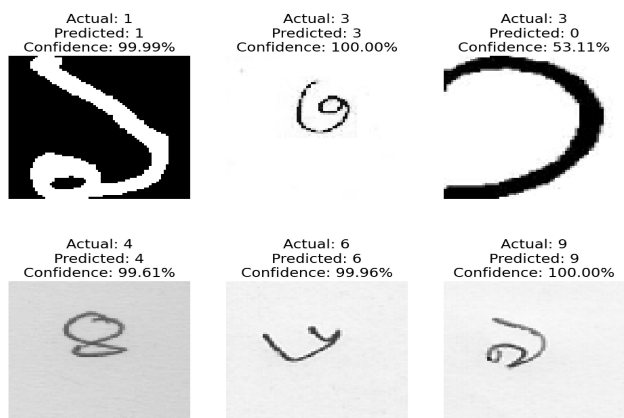

Fig. 5: Digit Predictions showing High Confidence for Correct and Low Confidence for Incorrect Cases.

V. Conclusion

This study demonstrates the successful application of the InceptionResNet model for Bengali handwritten digit recognition. The experimental results indicate that the fine-tuned InceptionResNet model performs effectively in recognizing Bengali handwritten digits. By fine-tuning a pretrained model and employing robust preprocessing techniques, we achieved high classification accuracy. However, some misclassifications occurred, particularly among visually similar digits.

VI. Future Enhancement

The methodologies and insights gained from this study can be leveraged to advance handwritten digit recognition in other languages as well. Researchers can explore adapting the proposed approach to languages with similar script characteristics. Additionally, extending the model to recognize characters beyond digits, such as alphabets or symbols, could broaden its applicability across various language scripts. Furthermore, collaborative multilingual datasets and benchmarks aid in building robust models for recognizing handwritten characters across various languages.

Acknowledgment

We sincerely thank the creators of the NumtaDB dataset for providing a valuable resource for Bengali handwritten digit recognition. We are also grateful to Google Colab for computational resources that enabled efficient training and experimentation. Special thanks to the open-source community for tools like TensorFlow and Keras, and to our mentors, peers, and families for their guidance, feedback, and unwavering support throughout this research.

References

[1] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. “Inception-v4, inception-ResNet and the impact of residual connections on learning,” In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI’17), AAAI Press, pp. 4278–4284, 2017.

[2] Ashikur Rahman, A. B. M., Hasan, Md. Bakhtiar, Ahmed, Sabbir, Ahmed, Tasnim, Ashmafee, Md. Hamjajul, Kabir, Mohammad Ridwan, and Kabir, Md. Hasanul, “Two Decades of Bengali Handwritten Digit Recognition: A Survey,” IEEE Access, vol. 10, pp. 92597-92632, 2022.

[3] A. Choudhury, H. S. Rana and T. Bhowmik, “Handwritten Bengali Numeral Recognition using HOG Based Feature Extraction Algorithm,” 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2018, pp. 687-690, doi: 10.1109/SPIN.2018.8474215.

[4] T. Hassan and H. A. Khan, “Handwritten Bangla numeral recognition using Local Binary Pattern,” 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Savar, Bangladesh, 2015, pp. 1-4, doi: 10.1109/ICEEICT.2015.7307371.

[5] A. Shawon, M. Jamil-Ur Rahman, F. Mahmud and M. M. Arefin Zaman, “Bangla Handwritten Digit Recognition Using Deep CNN for Large and Unbiased Dataset,” 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554900.

[6] R. Noor, K. Mejbaul Islam and M. J. Rahimi, “Handwritten Bangla Numeral Recognition Using Ensembling of Convolutional Neural Network,” 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2018, pp. 1-6, doi: 10.1109/ICCITECHN.2018.8631944.

[7] H. Zunair, N. Mohammed and S. Momen, “Unconventional Wisdom: A New Transfer Learning Approach Applied to Bengali Numeral Classification,” 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554435.

[8] Alam, Samiul, Tahsin Reasat, Rashed Mohammad Doha and Ahmed Imtiaz Humayun. “NumtaDB - Assembled Bengali Handwritten Digits.” ArXiv abs/1806.02452 (2018): n. pag.