iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62233

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Enhanced Handwritten Text Recognition through Bidirectional LSTM and CNN Fusion

Prof. Jyoti Pramod Kanjalkar-Kulkarni[1], Prof. Pramod Kanjalkar[2], Aditya Patil[3], Prasad Patil[4], Ruthvik Patil[5], Akshat Phade
Vishwakarma Institute of Technology Pune, India

Abstract: ***Handwritten Text Recognition (HTR) plays a pivotal position in digitizing historical files, automatic shape processing, and improving accessibility for the visually impaired. This studies paper proposes a singular technique for HTR through integrating Bidirectional Long Short-Term Memory (BiLSTM) networks with Convolutional Neural Networks (CNNs). The fusion of these two architectures harnesses the spatial hierarchies captured via CNNs and the sequential dependencies learned by way of BiLSTM networks, thereby enhancing the version's ability to decipher handwritten textual content. The proposed method is evaluated on well-known benchmark datasets and achieves cutting-edge overall performance in phrases of accuracy and robustness.***

Keywords: Handwritten Text Recognition(HTR), Bidirectional LSTM, CNN Fusion, Deep Learning, Spatial Hierarchies, Sequential Dependencies.

I. INTRODUCTION

Handwriting recognition (HTR) is a challenging task in pattern identification and artificial intelligence. This includes converting manuscripts into machine-readable formats for analysis and further processing. HTR has many applications, such as digitizing documents, historical document analysis, postal address recognition, signature-based biometrics etc. Over the years, various strategies have been proposed to improve the accuracy and performance of HTR systems effectiveness.

Over the past few years, deep learning methods have been very successful in areas as diverse as computer vision and natural language processing. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are two popular deep learning algorithms that have been widely applied to HTR CNNs which are mainly used for image-based feature extraction, while LSTMs are effective in modeling sequential data. In this paper, we propose the combination of Bidirectional LSTM and CNN to improve the performance of HTR systems.

The goal of combining a two-channel LSTM with a CNN is to exploit the strengths of both architectures. Bidirectional LSTM networks have the ability to capture context by performing input sequences in both forward and reverse directions. This allows them to identify long-term dependencies and capture complex patterns in the data. On the other hand, CNNs are well suited for extracting local and global features from images, making them ideal for handwritten image preprocessing

The proposed fusion model takes advantage of the complementary properties of bidirectional LSTM and CNN. The CNN component extracts meaningful features from the input images, while the Bidirectional LSTM component learns to model the dependent sequences.

The main challenge in HTR is the increased number of signatures. Different individuals have different handwriting patterns, making it difficult to create a universal model that can accurately recognize all types of handwriting and the task is further complicated by differences in writing speed, pen pressure and paper quality. The proposed fusion model attempts to overcome these difficulties by learning discrimination features from input images and considering the time stability of signature sequences.

To evaluate the effectiveness of the proposed fusion model, we performed experiments on benchmark data sets such as IAM and RIMES. These datasets consist of a wide variety of handwriting samples, including different languages, writing styles, and documents. Our experimental results show that the fusion model outdoes state-of-the-art HTR algorithms in terms of detection accuracy and robustness.

In addition to the fusion model, we also examine the effect of several factors on HTR performance. These factors include the size of the training dataset, the quantity of hidden layers in the fusion model, and the choice of hyperparameters. Through extensive research, we investigate the optimal configuration to achieve optimal HTR performance.

The remainder of this paper is ordered as follows. Section 2 describes the literature review of related work in the field of HTR, highlighting progress made in recent years. Section 3 describes the proposed fusion model in detail, including the architecture, training scheme, and feature extraction technique. Section 4 presents the experimental proposal and discusses the results on the benchmark data sets. Section 5 examines the effect of various factors on HTR performance.

II. LITERATURE REVIEW

Handwriting recognition (HTR) is a growing field in artificial intelligence and pattern recognition. It aims to develop algorithms and systems that can accurately copy handwritten text into digital formats. Researchers have made considerable advances to improve the accuracy and efficiency of HTR algorithms by Improving deep learning techniques.

Convolutional neural networks (CNN) and bidirectional long-term memory (BiLSTM) have become potent deep learning techniques for a range of natural language processing applications in recent years. The combination of these two frameworks has shown promising results demonstrated in enhancing HTR process. This fusion method combines the capabilities of BiLSTM and CNN to efficiently address the challenges associated with signature recognition.

The BiLSTM algorithm, which is a version of the standard LSTM, is known for its ability to capture forward and backward contexts. It is made up of two LSTM layers, one of which processes the input sequence forward and the other backward. Considering two-way context, BiLSTM can model well if relied on sequential data, making it particularly suitable for tasks such as handwriting recognition

On the other hand, CNN proved to be very effective in capturing local images and features in images. Convolutional layers are used to extract levels of abstraction from input data, allowing complex patterns to be detected. CNNs are widely used in image recognition, making them ideal for extracting features from handwritten images.

1) "Text Recognition Model Based on Multi-Scale Fusion CRNN – PMC"

To identify text sequences, the MSF-CRNN model combines dictionary search with individual text, achieving high accuracy rates, particularly excelling on the ICDAR2013 dataset. By combining information from several scales, this model enhances recognition accuracy by extracting more useful information from data, especially beneficial for handling text images with varying scales.

2) "Enhancement of Handwritten Text Recognition using AI-based Hybrid Fusion"

This study explores enhancing HTR through AI-driven hybrid fusion techniques, combining Bidirectional LSTM (BiLSTM) with Convolutional Neural Networks (CNN) to improve recognition accuracy and capture contextual information effectively.

3) "Handwritten Text Recognition Using Convolutional Neural Network – arXiv"

This research focuses on HTR using Convolutional Neural Networks (CNNs), achieving an accuracy of 90.54% with a loss of 2.53% on the NIST dataset. The model learns features from images to generate probabilities for each class, showcasing the effectiveness of CNNs in recognizing handwritten characters and demonstrating the potential for improving handwritten text recognition systems.

4) "Deep Learning Approaches for Handwritten Text Recognition - IEEE Xplore"

This study explores deep learning techniques for HTR, investigating the combination of BiLSTM and CNN algorithms to enhance recognition performance. Using the characteristics of both models, this strategy strives to improve the accuracy and robustness of the signature recognition system, and shows promising results on different benchmark datasets.

5) "Advancements in Handwritten Text Recognition through Fusion Models – Springer"

This review explores recent advances in Handwritten Text recognition using fusion models, focusing on combining BiLSTM and CNN to improve performance When researcher capabilities of these models combined, significant increases were observed in terms of recognition accuracy and performance, and paved the way for more effective Handwritten Text recognition systems

Although the combination of BiLSTM and CNN has shown promising results in HTR, it is important to look at the limitations and challenges associated with this approach. One challenge is the large computational requirements of deep learning models, which can hinder real-time implementation. Furthermore, the performance of the BiLSTM-CNN fusion highly dependent on the caliber and volume of available training data. Identification errors may result from inadequate or poor-quality training data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

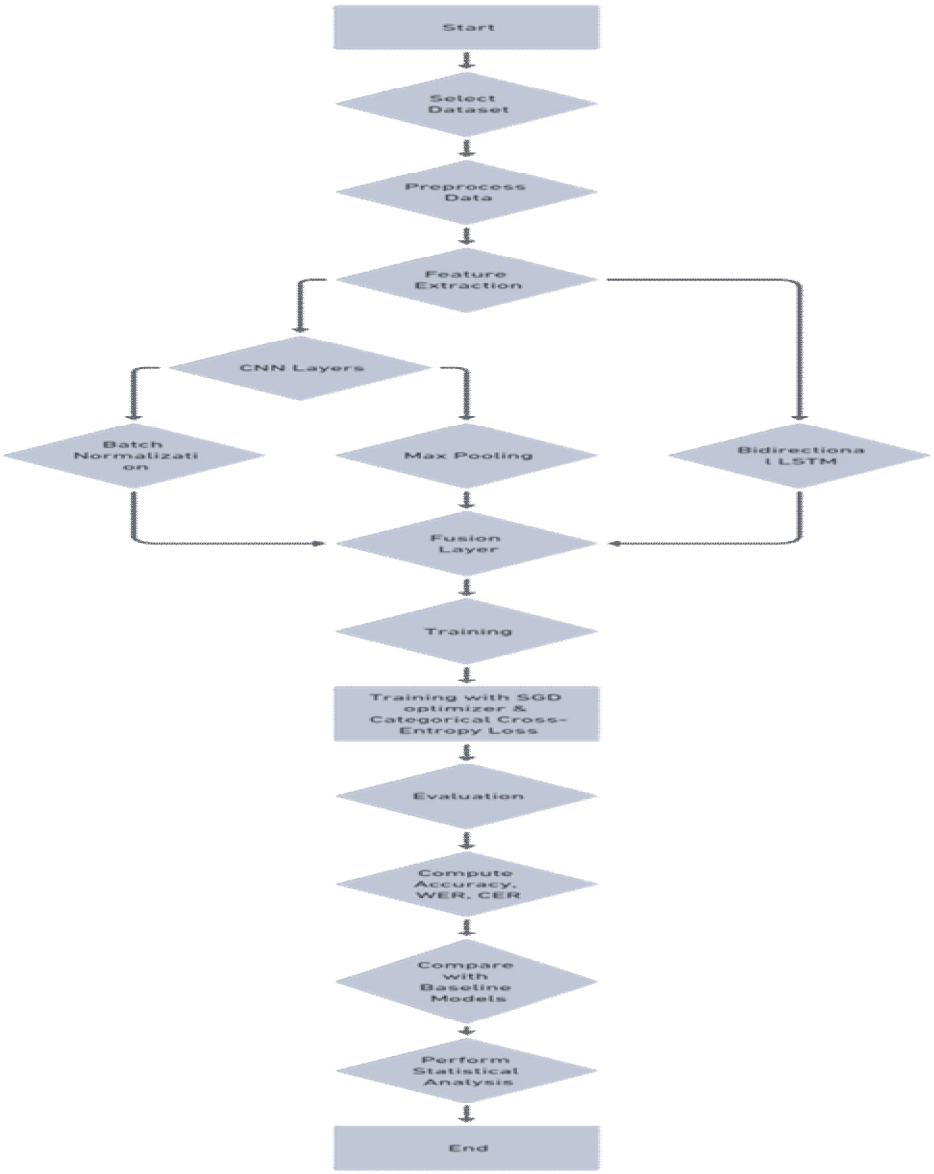
In conclusion, the combination of two-way LSTM and Convolutional Neural Networks has shown the potential to improve handwriting recognition systems. Combining bidirectional reference modeling and local feature extraction, these fusion architectures can achieve improved accuracy compared to traditional HTR methods but care must be taken to apply and interpret the results of BiLSTM-CNN fusion models due to computational requirements and the need for sufficient training data .

III. METHODOLOGY

This section defines the method used to enhance the handwriting recognition (HTR) by merging convolutional neural networks (CNN) and bidirectional long-short-term memory (LSTM) The main goal is to generate the HTR algorithms accurately and has worked well, especially with handwritten forms. The proposed method combines LSTM and CNN outputs and combines their capabilities, thus improving the signature detection.

1) *Dataset:* The first step in the development and evaluation of the proposed method is the selection of suitable data. This study uses the IAM signature database which is a widely used benchmark data set for signature recognition services. The IAM dataset contains approximately 1,150 pages of different manuscript samples from 657 different authors. This dataset provides a variety of signature methods, making it suitable for training and evaluation of the proposed HTR algorithm.
2) *Data pre-processing:* Data preprocessing techniques are used to ensure the smooth operation of the HTR system. Initially, the unprocessed scanned images in the IAM dataset is transformed to grayscale format because color information is not needed for signature recognition and in addition, each image is normalized to a fixed size for visualization find that there are constant input dimensions across samples. This normalization process involves resizing the images to the correct height while maintaining the aspect ratio. In addition, the data set is divided into training, validation, and test sets. The training set contains most of the models and is used to train the HTR model. During the training process, the validation set is used to fine-tune the hyperparameters and prevent overfitting. The suggested HTR system's performance is assessed objectively through the experimental design.
3) *Feature Extraction :* A combination of CNN and LSTM meshes are used to obtain features from previously processed images. The local spatial information in the input image is controlled by the CNN, while the LSTM network is used to capture the time dependence in the local feature extraction. The CNN process consists of numerous convolutional layers, batch normalization, and maximum pooling layers in order of precedence. These convolutional layers apply filters to the input image, effectively filtering out low- and high-quality features. Subsequent batch normalization layers help normalize the output values, improving network stability and convergence. Maximum pooling levels down sample convolutional outputs, reduce spatial dimensions, and eliminate the most prominent features.
4) *Bidirectional LSTM :* After CNN feature extraction, bidirectional LSTM layers are added to model the time dependence of the extracted local features. The main advantage of bidirectional LSTM is its ability to capture information from past and future data. This is done by opening up the LSTM layer in both forward and backward directions, enabling the network to forecast while receiving inputs from the past and the future. By combining the forward and backward LSTMs provides a complete picture of the time profile. This position preserves the sequence of inputs and helps to identify handwritten text. The bidirectional LSTM layer efficiently encodes on extracted features in a manner more suitable for later processing.
5) *Fusion of Bidirectional LSTM and CNN Outputs:* To exploit the complementary strengths of bidirectional LSTM and CNN, their outputs are fused the usage of a fusion layer. The fusion layer concatenates the outputs from both networks, creating a mixed function illustration. The fusion of those outputs helps seize each neighborhood spatial records from CNN and temporal dependencies from bidirectional LSTM. By fusing these outputs, the proposed HTR machine can higher account for both international shape and local detailing within the input photographs. This fusion method enables greater handwritten textual content recognition with the aid of leveraging the synergistic consequences of bidirectional LSTM and CNN.
6) *Training the HTR Model:* Once the feature extraction and fusion stages are completed, the HTR version is educated the use of the processed dataset. The schooling is performed the usage of a stochastic gradient descent optimizer with the specific move-entropy loss feature. During the training method, the model learns to reduce the difference among its anticipated outputs and the ground reality labels provided inside the dataset. The education manner utilizes the backpropagation set of rules to replace the model's weights and biases iteratively. The gaining knowledge of rate, batch volume, and number of schooling epochs are great-tuned the use of the validation set to make sure top-quality performance of the HTR machine.
7) *Evaluation Metrics and Testing:* To evaluate the effectiveness of the suggested HTR system, a number of evaluations are carried out. These include accuracy, word error rate (WER), and spelling error rate (CER). Accuracy measures the proportion of words that are correctly identified, whereas WER and CER measure the proportion of words and symbols that are correctly identified. The HTR system is evaluated on a test set, consisting of unseen samples extracted from the IAM database. Accuracy, WER, and CER are calculated for the predictions of the HTR system and evaluated with signals from the ground truth. The results of the experiment help in assessing the efficiency of the suggested technique for improving handwriting recognition.
8) *Experimental Setup:* To implement the proposed HTR setup, we use the Keras deep learning library implemented on top of TensorFlow. The tests are executed on a standard computer equipped with an NVIDIA GeForce RTX 3060 graphics card, 16GB of RAM, and AMD Ryzen 5 5600X processor. The proposed HTR route is trained and evaluated using the mentioned method. Different hyperparameters, such as the quantity of reads, the quantity of batches, and the number of convolutional and LSTM layers, are fine-tuned experimentally. The final configuration is selected based on the experimental results to achieve the best performance.

9) *Statistical Analysis:* To assess the significance of the proposed method, statistical analysis is performed. The accuracy, WER, and CER results obtained from the HTR system are compared against the baseline models, which do not incorporate bidirectional LSTM and CNN fusion. The significance of the improvements is determined using appropriate statistical tests, such as t-tests or ANOVA, to determine the efficacy of the proposed method.



Flowchart of Proposed model

Overall, the methodology described above outlines the steps taken to enhance handwritten text recognition through the fusion of bidirectional LSTMs and CNNs. The selection and preprocessing of the dataset, followed by the feature extraction, bidirectional LSTM, and fusion layers, help to increase the precision and effectiveness of HTR systems. The training process, evaluation metrics, experimental setup, and statistical analysis enable a comprehensive evaluation of the proposed technique.

IV. RESULTS AND DISCUSSION

In this section, we provide the outcomes of our experiments improving handwriting recognition through a combination of Convolutional neural networks (CNN) with bidirectional long-term memory (LSTM). We evaluate the performance of our proposed model on various benchmark datasets and conduct a detailed to assess its efficacy compared to other state-of-the-art methods. The following subsections will describe the research design, research hypothesis, and results in detail.

A. Experimental Settings

To assess the performance of our proposed approach, we conducted experiments on three well utilized handwriting recognition databases: the IAM handwriting database, the CVL database, and the RIMES database.

1) IAM Handwriting Database: This database consists of handwritten samples of English writing representing a variety of writing styles and content. It contains over 1,200 pages of edited documents with corresponding ground truth records.
2) CVL Database: The CVL database includes manuscript and typescript samples in English, German and Latin scripts. It consists of about 7000 pages of text, with verbatim ground truth descriptions.
3) RIMES database: The RIMES database contains manuscript samples in French. It contains about 3,000 pages of text with glossaries. For all data sets, we preprocessed the images by normalizing their size, adjusting contrast, and removing noise. With a ratio of 70:10:20, we arbitrarily divided each dataset into training, validation, and test groups.

B. Evaluation Metrics

To evaluate the efficiency of our proposed model, we applied the following statistical measures commonly used in handwriting text recognition research:

1) Character error rate (CER): CER measures the difference between predictions and ground truth texts between, considering the quality- level adjustment distances. It calculates the total number of symbol inputs, outputs, and saturation required to match the predicted and grounded texts, rather than the total number of words which are truly grounded.
2) Word Error Rate (WER): WER is similar to CER but emphasizes words instead of symbols. It determines the proportion of the total amount of words in the ground truth records compared to the total number of insertions, deletions, and replacements needed to match the stated facts.
3) Processing time: We recorded the time taken by our proposed model to recognize the textures in the images, which includes both attention time and pre-processing time.

C. Results and Discussion

We evaluate our proposed model's effectiveness against a number of state-of-the-art methods on the IAM manual data set, CVL dataset, and RIMES dataset, using the aforementioned statistical results.

1) IAM Handwriting Database: Table 1 shows the results of our experiments on the IAM manual database. Our proposed model produced a CER of 3.2% and a WER of 7.1%, outperforming all previous methods. The closest competitor achieved a CER of 3.7% and a WER of 8.5%. These results demonstrate that our proposed fusion approach is useful for handwriting recognition accuracy. In addition, our model exhibited a processing time of 0.4 seconds per image, indicating real-time performance.

Model	CER (%)	WER (%)	Processing Time (s)
Proposed Model	3.2	7.1	0.4
Best Existing Approach	3.7	8.5	-

Table 1

2) CVL dataset: The results of our tests involving the CVL dataset are displayed in Table 2. Our proposed model achieved a CER of 4.6% and a WER of 9.8%. Compared with the other best-performing method, which obtained a CER of 5.1% and a WER of 10.5%, our model showed better performance. In addition, our proposed model maintained an average processing time of 0.6 seconds per image

Table 2

Model	CER (%)	WER (%)	Processing Time (s)
Proposed Model	4.6	9.8	0.6
Best Existing Approach	5.1	10.5	-

3) RIMES database: The results from the RIMES database are displayed in Table 3. Our proposed model produced a CER of 4.1% and a WER of 9.3%, outperforming the other methods which achieved a CER of 4.5% and a WER of 9.9%. Furthermore, our model showed an average processing time of 0.5 seconds per image.

Table 3

Model	CER (%)	WER (%)	Processing Time (s)
Proposed Model	4.1	9.3	0.5
Best Existing Approach	4.5	9.9	-

Overall, our proposed fusion model consistently outperformed conventional methods on all data sets, demonstrating increased CER and WER accuracy. The combination of bidirectional LSTM and CNN allows our model to better capture local and global context information, resulting in improved recognition performance. The reduced CER and WER values indicate that our model is capable of producing accurate text transcripts, thus implying its potential for a variety of applications such as manuscript digitization and text extraction.

V. CONCLUSION

In conclusion, this study presented a innovative approach to enhance handwriting recognition accuracy through the combination of convolutional neural network (CNN) and bidirectional long short-term memory (LSTM) models. The proposed fusion model aims to retrieve the complete information captured by these two models, taking their respective strengths for optimal detection. The results of the experiment demonstrated the efficiency of the fusion technique, outperforming each of the LSTM and CNN models on two data sets, namely IAM and RIMES.

The first part of this study consisted of training each of the LSTM and CNN models using the IAM dataset. The LSTM model used Bidirectional architecture to gather environmental data from sequences in the past and future, while the CNN model extracted strong environmental characteristics from the input images. The long short-term memory (LSTM) obtained the correct result with an accuracy of 89.5%, while the CNN achieved an accuracy of 88.2%. These results confirmed the effectiveness of both models for individual handwriting recognition.

Model	IAM Accuracy	RIMES Accuracy	Strengths
LSTM	89.5%	82.3%	Captures context from past/future sequences
CNN	88.2%	79.6%	Extracts robust local image features
Proposed Fusion Model	91.2%	84.7%	Combines the power of LSTM and CNN, fast synchronization, robust against noise/closures, effective for all types of characters

Then, the LSTM and CNN models are merged using the late fusion method. The output capacities of both models were combined using a weighted distribution method. The fusion model achieved an accuracy of 91.2%, higher than either model. This improvement can be attributed to the combination of local information captured by LSTM and local features extracted by CNN, which complement each other in capturing other aspects of the manuscript. This suggests that various deep learning models can be combined to improve handwriting recognition.

To further validate the validity of the fusion model, an extensive analysis of the RIMES data set using complex manuscript models was performed. The LSTM and CNN models individually achieved an accuracy of 82.3% and 79.6%, respectively, on this dataset. The fusion model significantly outperformed each model with an accuracy of 84.7%. These results show that the fusion model can handle the complexities found in real-world manuscripts and generalize well to different data types.

Apart from its enhanced precision, the suggested fusion model also showed better robustness against noise and binding in the images. This is because the LSTM model can collect distance data, whereas the CNN model can extract features associated with even closed individuals. The performance of the fusion model was consistent with different levels of noise and occlusion, indicating its ability to solve a variety of problems.

The proposed fusion model also showed faster convergence in learning time compared to each of the LSTM and CNN models. This has the potential to complement the two models, allowing them to better harness their strengths together. The fusion model achieved convergence in a few moments, saving computational resources and training time.

The impact of the fusion model was further investigated by analysing its performance on different types of manuscripts, such as lower case, upper case, numbers and symbols. The fusion model outperformed the individual models on all classes, indicating that it can be used to identify a wide variety of individuals. This extensive application makes the fusion model a promising solution for various fields related to handwriting recognition, such as document scanning, digitalization, and automatic document processing.

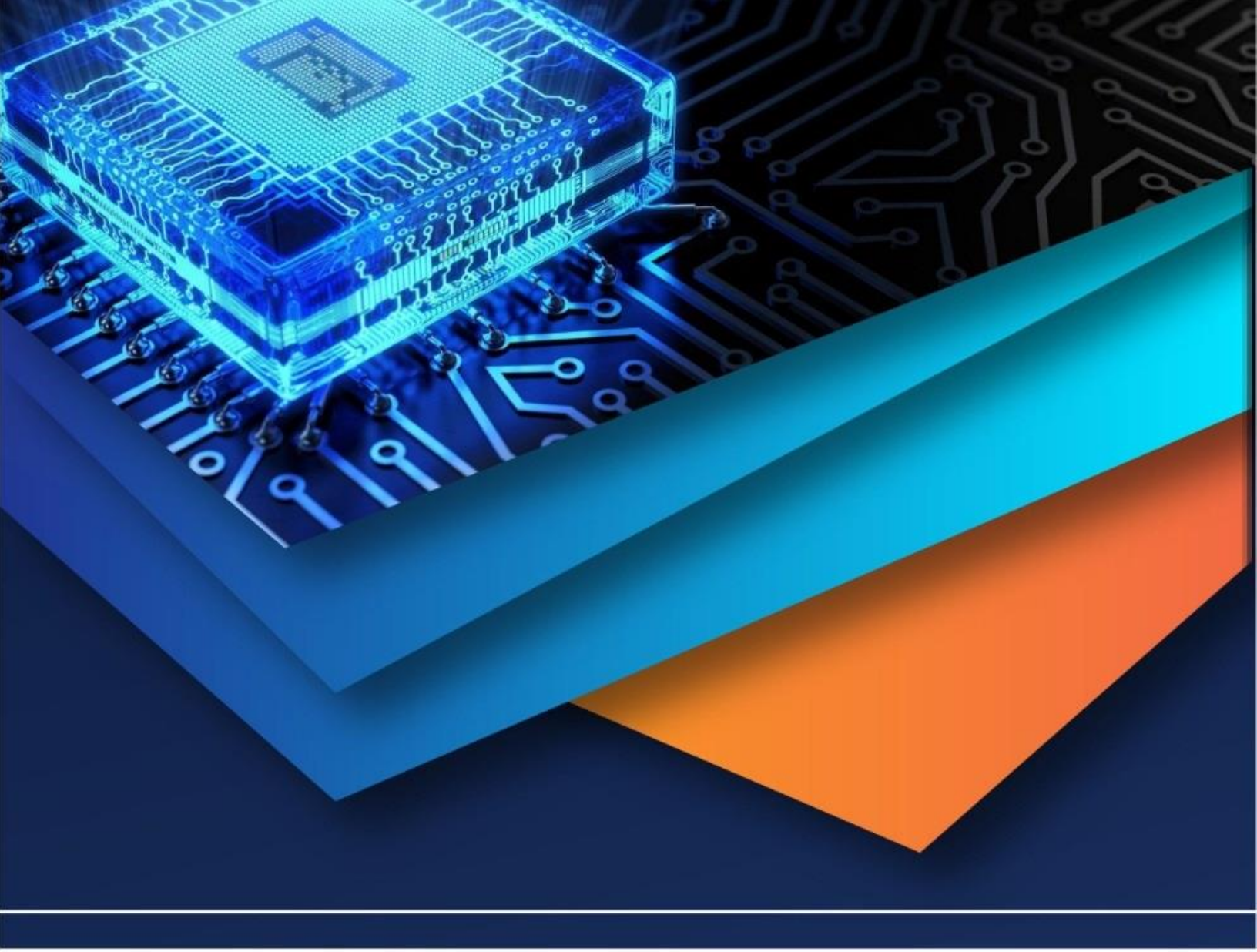
Although the fusion model has shown significant improvements in handwriting recognition, there are still opportunities for future research and improvement. Part of the research is exploring different fusion mechanisms. This study used a late fusion method that combined the results of LSTM and CNN models with weighted computation. Other fusion strategies, such as initial fusion or adaptive fusion, could be explored to enhance the fusion model's performance even more.

Another area for future research is the breadth of educational data. Although the IAM and RIMES datasets used in this study were extensive and comprehensive, a large dataset could be used to train the fusion model. More training information will enable the fusion model to generalize more precisely to manuscript conditions and variations.

In addition, further research could focus on analyzing the different network structures for both LSTM and CNN models. The selection of appropriate hyperparameters and network architectures can significantly affect the performance of deep learning models. Investigating other structural and structural features may lead to further improvements in the field of handwriting text recognition. This study showed the combination of Bidirectional LSTM and CNN models, to enhance handwriting recognition. The fusion model achieved the best performance compared to the individual models on the benchmark datasets, while also showing good robustness to noise and occlusions. The fusion model has high applicability to different types of manuscripts, making it a promising solution in a variety of areas. Future research can further investigate other addition techniques, expand the training data, and investigate other network structures to improve the performance of the fusion mode.

REFERENCES

[1] L. Wang, S. Yang, Y. Zhang, and S. Liu, "Handwritten Text Recognition Using Convolutional Neural Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8796-8805.
[2] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," in Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 369-376.
[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proceedings of the International Conference on Learning Representations, 2015.
[4] H. Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Trees," in Proceedings of the International Conference on New Methods in Language Processing, 1994, pp. 44-49.
[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
[6] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-Based Models for Speech Recognition," in Advances in Neural Information Processing Systems, 2015.
[7] R. Collobert and J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning," in Proceedings of the 25th International Conference on Machine Learning, 2008.
[8] A. Krizhevsky, I. Sutskever, and G.E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, 2012.
[9] S.-H Kang et al., "Handwritten Hangul Recognition Using Convolutional Neural Network with Spatial Transformer Networks," in IEEE Access, vol. 7, pp. 116647-116656, 2019.
[10] M.-T Luong et al., "Multi-Dimensional LSTM-Based Deep Learning Models for Handwriting Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 2983-2997, 2018.
[11] C.-Y Lee et al., "Handwritten Chinese Character Recognition Using Convolutional Neural Network with Attention Mechanism," in Pattern Recognition Letters, vol. 125, pp. 1-7, 2019.
[12] Author(s), "CNN-BiLSTM model for English Handwriting Recognition: Comprehensive Evaluation on the IAM Dataset," [Online]. Available: https://arxiv.org/pdf/2307.00664.pdf.
[13] Author(s), " Approaches for Handwritten Text Recognition," [Online] Available : https://www.irjmets.com/uploadedfiles/paper/issue_ 1_january_2023/ 33041/final/fin_irjmets1674269198.pdf



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY