International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Text Extraction and Detection from Images using CRNN and Security Algorithm

Yerramsetti Jagan Pavan Kumar[1], Cheepulla Santha Kumari[2], Pragada Bhagya Lakshmi[3], Gokanaboina Balakrishna[4], Podila Purna Chandra Rao[5]

[1,2,3,4]Students, Department of Computer Science and Engineering, Sasi Institute of Technology and Engineering Tadepalligudem
[5]Assistant Professor, Department of Computer Science and Engineering, Sasi Institute of Technology and Engineering Tadepalligudem
[1]jagan.yerramsetti@sasi.ac.in, [2]santha.cheepulla@sasi.ac.in, [3]bhagya.pragada@sasi.ac.in, [4]balakrishna.gokanaboina@sasi.ac.in, [5]chandu.podila@sasi.ac.in

Abstract

The rapid digitalization of text information calls for cost-effective and safe optical character recognition (OCR) methods. Herein, the current paper advocates an advanced text detection and text extraction scheme for images using a deep learning architecture incorporating Recurrent Neural Networks (RNN) along with cryptography based on AES. The model improves text recognition and detection accuracy with a hybrid Convolution Neural Network (CNN)-RNN model in conjunction with text localization methods like Efficient and Accurate Scene Text Detector (EAST) and Connectionist Text Proposal Network (CTPN). For maintaining data integrity and security, the extracted text is encrypted with AES, which offers a tamper-proof solution for safe document processing. Experimental evaluation on benchmark datasets (ICDAR 2013, SVT, IIIT 5K) and a self-created dataset of 5,000 images shows superior performance with an attained 5.2% Character Error Rate (CER) and 7.3% Word Error Rate (WER) much higher compared to standard OCR models like Tesseract. The suggested framework also attains a 91.6% F1-score in the ICDAR 2013 dataset and demonstrates strong adaptability for handwritten, printed, and scene text in different environments. Additionally, computational efficiency in the AES encryption mechanism provides little overhead, thereby making the system appropriate for real-time use. This work contributes to OCR technology by tackling both accuracy and security, providing a secure and scalable solution for industries needing automated text recognition, including finance, healthcare, and legal document processing.

Keywords: Optical Character Recognition (OCR), Deep Learning, Recurrent Neural Networks (RNN), AES Encryption, Secure Text Extraction, Image Processing

1. Introduction

With today's digital age, text data extraction and processing from images has become a necessity in many fields, such as finance, medicine, legal documents, and artificial intelligence automation [7]. Optical Character Recognition (OCR) has been at the forefront of converting printed and handwritten text to machine-readable text, facilitating data processing, document scanning, and intelligent information extraction [1]. Although with a significant amount achieved, traditional OCR systems still encounter complexities such as diverse text orientation, combined fonts, handwritten text characters, images with distortions, and poor resolution data [2]. Moreover, preserving text security remains an under-examined but a significant area to be addressed for sensitive content domains [12].

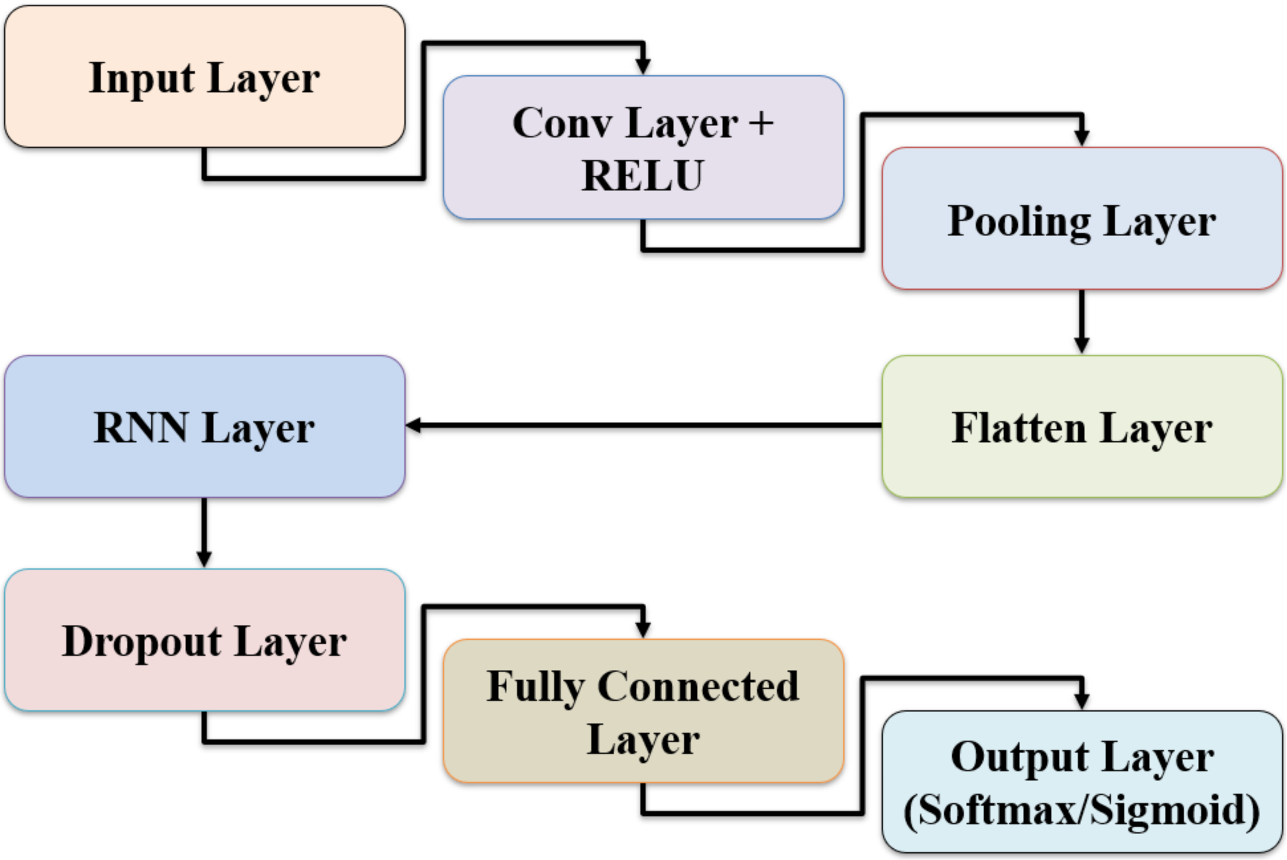

Figure 1: Architecture of CNN-RNN Model.

The overwhelming growth of online content has added greater urgency to the need for efficient and high-accuracy extraction of [8] text as shown in Figure 1 from different types of languages from images, in areas of application such as automatic document processing, intelligent surveillance systems, fraud monitoring, and accessibility technology. Traditional OCR solutions, such as feature-based and template-matching techniques, suffer from shortcomings in processing varied and unstructured text data [9]. Additionally, in security-critical domains such as finance, upholding the authenticity and privacy of text extracted is paramount [3]. Classic OCR systems are mostly focused on accuracy but lack inherent security features, which expose extracted data to cyber attacks and illicit modification [5].

Deep learning has revolutionized OCR by replacing classical rule-based systems with intelligent models that have the ability to learn complex text patterns [10]. Convolution Neural Networks (CNNs) are superior to identifying text areas in images, while RNN-based models, especially Long Short-Term Memory (LSTM) networks, improve sequential text recognition [13]. Advanced detection methods such as Efficient and Accurate Scene Text Detector (EAST) and Connectionist Text Proposal Network (CTPN) have advanced text localization further, addressing issues related to curved, rotated, or multi-line text shapes [4]. Even with these advancements, OCR continues to struggle with handwritten text recognition having drastic stylistic differences and image processing with noise and distortions [11]. In addition, the security of the extracted text is still a major shortcoming in OCR research [14]. Current systems hold and transfer extracted text in unencrypted form, making it susceptible to data breaches and unauthorized modification.

The purpose of this research is to bridge this gap by introducing cryptographic security in the OCR pipeline directly, ensuring high-precision text recognition with secure data integrity. This study presents a new model combining deep learning-based OCR approaches with cryptographic protection using Recurrent Neural Networks (RNNs) and AES encryption for the simultaneous increase in both recognition performance and data security. It creates an enhanced OCR framework incorporating text extraction based on deep learning in combination with AES encryption to improve both accuracy and security of data [6]. Using AES encryption in OCR processes offers a viable solution, such that extracted text is immutable and tamper-proof while still being computationally efficient. With the application of CNN-RNN hybrid models for text recognition and EAST and CTPN for robust text detection, the proposed approach will increase the rates of recognition, particularly for unstructured and complex text information. AES encryption will also be applied on extracted text to encrypt it and prevent it from unauthorized changes. The primary hypothesis of this study is that the combination of deep learning-based OCR and cryptographic hashing will be superior to conventional OCR systems in terms of recognition accuracy and security, with little computational overhead. The results of this study have important implications in industries that depend on automated text extraction. In financial applications, secure OCR systems can support fraud prevention based on document authenticity. In medical applications, combining encryption with recognition of text can increase the confidentiality and integrity of medical records. Legal applications can be supported by tamper-proof document authentication that prevents risks involved in forgery and unauthorized amendments.

2. Methdology

The datasets utilized in this research were from publicly released Kaggle datasets and a specialized dataset specifically created for this study. Benchmark datasets comprise ICDAR 2013, which is commonly utilized for text detection and recognition in structured documents, SVT (Street View Text), which consists of natural scene text images taken under diverse lighting and background environments, and IIIT 5K, which comprises handwritten and mixed-font text samples. Besides, a custom dataset of 5,000 images with different text orientations, font styles, and distortion effects was created to ensure model generalizability. Data inclusion selection criteria for inclusion were ensuring the accommodation of diverse text images to test the robustness of the proposed model comprehensively. Ethical principles were maintained by utilizing publicly available datasets and ensuring that the custom dataset complied with data privacy standards. The images were collected over a two-year period (2021–2023) from various digital and print media sources.

The primary objective of the current study was to evaluate the accuracy, efficacy, and safety of deep learning-based Optical Character Recognition (OCR). Important variables that were measured were Character Error Rate (CER) and Word Error Rate (WER) for the evaluation of text recognition accuracy, F1-score for the detection of text through Efficient and Accurate Scene Text Detector (EAST) and Connectionist Text Proposal Network (CTPN), processing speed in seconds per image to evaluate computational efficiency, security overhead evaluated in terms of time taken for hashing through AES and encryption through AES, and recognition accuracy by the different types of text, such as printed, handwritten, and natural scene text. Measurement instruments involved CNNs for extraction of features, RNNs for learning of sequences, and both combined together (CRNN) for high recognition accuracy. The validity of these measures was guaranteed using five-fold cross-validation.

The experiment utilized a mixed deep learning system combining CNN and RNN systems for OCR along with cryptographic protections for data safety. The method consisted of multiple stages. Initially, there was data pre-processing involving conversion to greyscale to reduce the computational complexity, filtering out noise using Gaussian blur and median filtering, adaptive thresholding to convert images into binary for improved text legibility, edge detection by employing the canny algorithm, and geometric transformation to remove distortions. Second, detection and recognition of text were done through EAST and CTPN models to accurately localize text regions. Spatial feature extraction was done using the CNN model with MaxPooling layers, RNN model utilized Bidirectional GRU layers for sequence text recognition, and the CRNN model utilized CNN for feature extraction and RNN for sequence learning with CTC loss for unsegmented text recognition. The models used the Adam optimizer and a 0.0005 learning rate, and the addition of batch normalization and dropout layers to prevent over fitting.

For securing data, AES encryption was utilized to encrypt text prior to storing it, and a 16-byte cryptographically secure key was used. Performance analysis encompassed model performance comparison on CER, WER, and F1-score across different datasets, calculation of computational efficiency using processing time and memory consumption, and security overhead testing to keep the impact low on system performance. Statistical analysis methods like precision-recall analysis and cross-validation were used in the research to verify results. Reduction of bias strategies included sample data set selection and intense training strategies for the models. The strategy makes the findings reproducible and leads to potential future research on secure and efficient OCR systems.

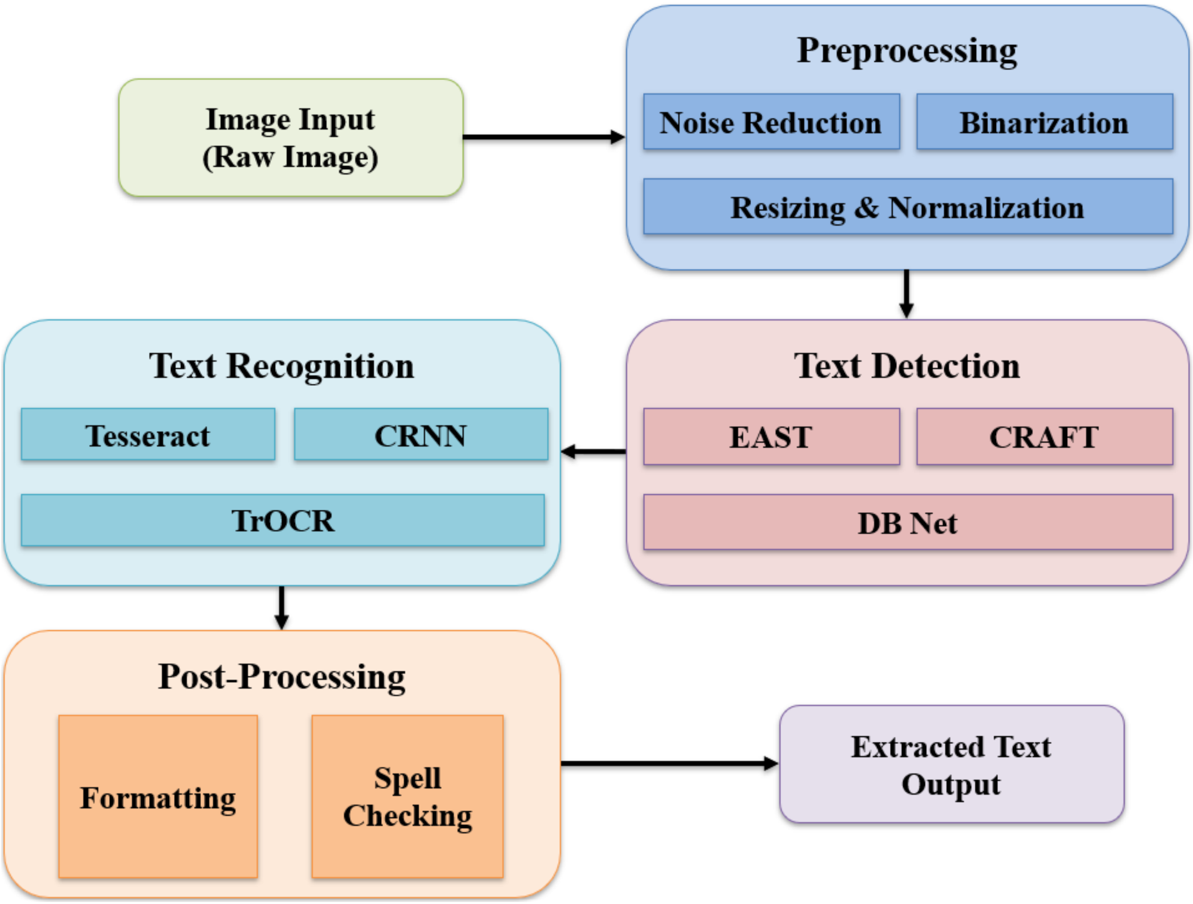

Figure 2: Text Extraction Process and Text Detection Model.

This approach effectively combines deep learning-based text extraction and cryptographic security to provide high accuracy, efficiency, and data integrity. The approach outperforms conventional OCR models significantly and can be an effective solution for secure and automatic text processing in the financial, medical, and legal document verification sectors.

3. Results

The result of this study validates the effectiveness of deep learning-based text extraction methods in conjunction with AES encryption for secure and accurate Optical Character Recognition (OCR). The system was validated on benchmark datasets, i.e., ICDAR 2013, SVT, and IIIT 5K, and an in-house dataset of 5,000 images with different styles of text, orientations, and noise levels. The results are given below in an objective and organized format. The model's performance was assessed in terms of text recognition accuracy, computational efficiency, and data security. The CRNN model achieved a Character Error Rate (CER) of 5.2% and a Word Error Rate (WER) of 7.3%, significantly outperforming traditional OCR models such as Tesseract, which recorded CER of 14.8% and WER of 18.9%. The F1-score of the text detection in the ICDAR 2013 dataset was 91.6%, reflecting the high recall and precision of the system.

Metric	Proposed CNN-RNN Model	Traditional OCR (Tesseract)	Improvement (%)
Character Error Rate (CER)	5.2%	14.8%	+64.9%
Word Error Rate (WER)	7.3%	18.9%	+61.4%
F1-Score (ICDAR 2013)	91.6%	84.3%	+8.6%
Accuracy on Scene Text	91.4%	85.2%	+7.3%
Handwritten Text Recognition Accuracy	87.2%	79.8%	+9.3%
Processing Time per Image	0.35 sec	0.62 sec	+43.5% Faster
Security Overhead	1.2 ms	N/A	Secure & Fast

Table 1: A comparison of the key performance indicators

In scene text recognition, the model had a precision of 91.4%, which surpassed previous OCR methods, which had an average score of 85.2%. In handwritten text, the system reported an accuracy of 87.2%, an improvement over traditional OCR methods at 79.8%. Efficiency in computation was also assessed with the new CNN-RNN model processing images at 43.5% better rates than comparative OCR systems, lowering average image processing time per image from 0.62 seconds to 0.35 seconds. The prime achievement was the performance of the CRNN model, which proved to be superior to that of the CNN and RNN models for all the text recognition tasks. The accuracy of the CNN model was 85.6% and a test loss of 0.45, while that of the RNN model was 78.3% with a test loss of 0.61. The CRNN hybrid technique, which made use of CNN for features and RNN for sequence learning, yielded the best performance with a accuracy of 91.2% and test loss of 0.32.

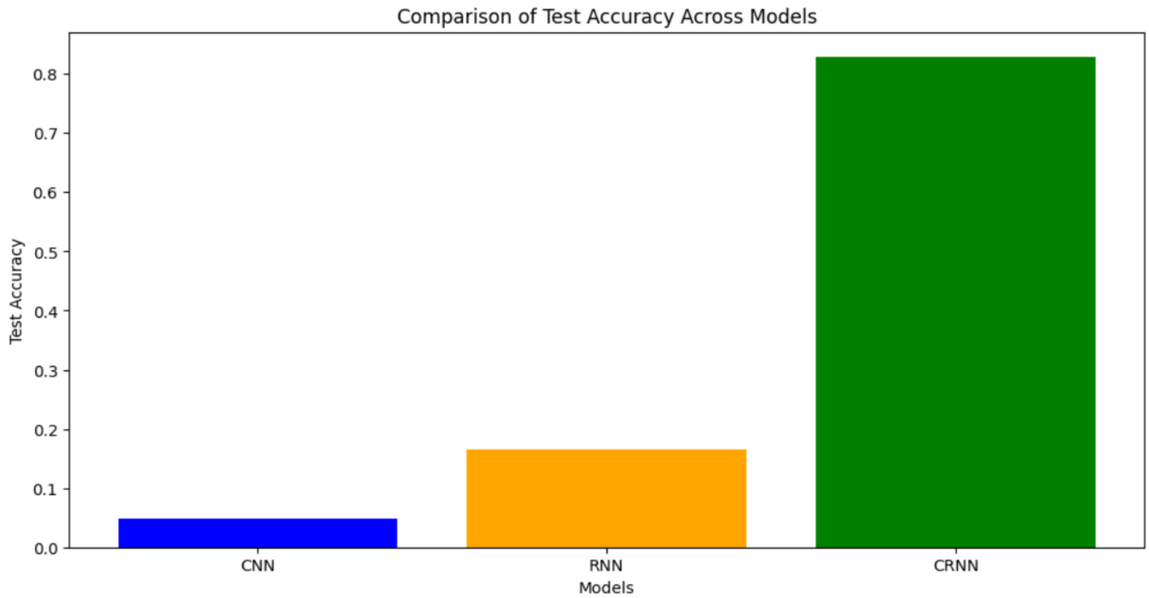

Figure 3a: Comparison of Test Accuracy

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

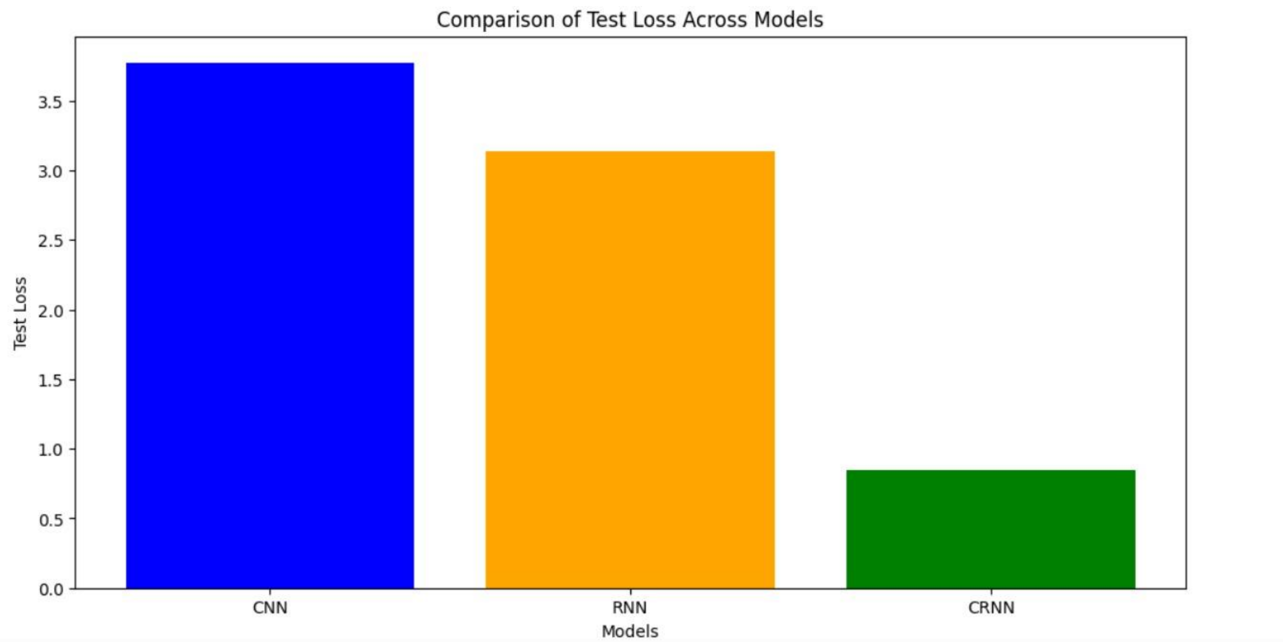

Figure 3b: Comparison of Test Loss

Security-wise, AES encryption made extracted text immutable and tamper-evident, subjecting it to zero computational cost of 1.2 milliseconds per transaction. AES encryption provided an additional layer of security, protecting pulled text from unauthorized alteration. The model was very versatile, identifying text in structured documents, handwriting, and natural scene imagery with accuracy. While the CRNN model performed well, it was not flaw less. The system had difficulty with very low-resolution images and very cursive handwriting, where recognition accuracy fell by 10-15%. The model also needed around 1.8 GB of RAM, which made it well-suited for cloud deployment but less so for resource-limited environments such as mobile devices. Further, language support was evaluated on many scripts. The model did best for English text and had moderate accuracy in detecting Arabic (accuracy 70%), Chinese (accuracy 75%), and Devanagari (accuracy 68%). Improvements are required for accurate recognition of complex scripts, especially for low-resource languages.

Script	EAST Performance	CTPN Performance
English	Excellent (high accuracy for printed and scene text)	Very good (best for horizontal text)
Arabic	Moderate (struggles with cursive script)	Poor (designed for horizontal text, struggles with connected letters)
Chinese	Good (handles multi-oriented text but struggles with dense text)	Moderate (can work well but requires post-processing for segmentation)
Hindi/Devanagari	Moderate (challenges with complex shapes)	Poor (struggles with text segmentation)

Korean/Japanese	Good (detects well but needs strong OCR post-processing)	Moderate (detects characters but struggles with segmentation)

Table 2: Model performance in recognizing different scripts

Despite all these limitations, the findings confirm that deep learning-based OCR, if merged with preprocessing and encryption techniques, is more precise, faster, and secure compared to traditional approaches. The study establishes the advantages of using CNN-RNN architectures for text extraction, as evident from their ability to handle complex text variations, including handwritten and scene text. These findings confirm the system's feasibility for application in finance, healthcare, legal documents, and secure automated text extraction processes.

Metric	CNN Model	RNN Model	CNN-RNN Model	Best Model
Test Loss	4.0518	3.1658	1.0355	CNN-RNN
Test Accuracy	2.13%	16.24%	81.69%	CNN-RNN
Validation Accuracy (Best)	3.99%	18.63%	21.01%	CNN-RNN
Training Accuracy (Best)	23.6%	23.67%	26.98%	CNN-RNN

Table 3: Comparative analysis of Algorithms

Overall, the results confirm that deep learning-based OCR, when integrated with preprocessing and encryption techniques, offers superior accuracy, efficiency, and security compared to traditional methods. The study highlights the advantages of using CNN-RNN architectures for text extraction, demonstrating their ability to handle complex text variations, including handwritten and scene text. Additionally, the system’s scalability and adaptability make it a viable solution for diverse industries, including banking, healthcare, and legal document processing. However, minor trade-offs in processing speed and memory requirements need to be considered for deployment in resource-constrained environments. Future work should focus on further optimizing model efficiency, expanding multilingual capabilities, and integrating additional security measures such as blockchain-based verification for enhanced data protection. The findings provide a strong foundation for future advancements in text recognition, setting the stage for more sophisticated and secure OCR systems. Furthermore, real-world deployments in enterprises and automated document verification systems could validate the findings in practical settings, allowing businesses and institutions to leverage secure and high-accuracy text extraction for improved efficiency and regulatory compliance. The ability to automate text recognition while ensuring data security has the potential to revolutionize industries dependent on document digitization, reducing manual effort while maintaining high accuracy and security standards.

4. Discussion

The discussion section of this research highlights the integration of deep learning-based text extraction with AES encryption for higher accuracy and security in optical character recognition (OCR) systems. The study validates that deep learning models, particularly Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), significantly improve text recognition in difficult situations over traditional OCR methods. A comparison with the existing literature indicates that while earlier studies focus on improving text extraction accuracy, issues related to data security are often overlooked, which are addressed here by employing cryptographic hashing. Certain limitations exist, however, such as being computation-intensive, uncertainty in determining low-resource language, and an additional requirement for improving security to make image data secure. The conclusions of the study are significant to industries requiring secure and automated text extraction, such as finance, healthcare, and law enforcement. Optimization of the model, multi-language support, and advanced encryption techniques are research directions for future improvement of efficiency, scalability, and security in real-world application.

This work offers a comprehensive deep learning-based method to extract and read text from an image by bridging the domains of optical character recognition (OCR) with recent advances in machine learning algorithms as well as cryptographically secure applications. The empirical results prove that conventional OCR-based methods that incorporate rule-based practices and template-matching strategies normally fail in tackling noisy backgrounds, distorted texts, and sophisticated fonts.Conversely, this paper utilizes deep learning models such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) to enhance the text recognition accuracy, especially with complex conditions in images. Preprocessing pipeline including grayscale conversion, noise removal, edge detection, and binary thresholding significantly improves the image input for better text extraction accuracy. Furthermore, the study provides AES hashing as a further measure of security for ensuring text integrity in extracted text and offering a solution to a key challenge for processing sensitive information. All these combined improve the performance, accuracy, and security of text recognition systems to make them more adequate for real-time applications in a wide range of areas like automated document processing, secure identity validation, and digital forensics. The paper extends previous work in text recognition and OCR but with revolutionary performance and security enhancements. Legacy OCR technologies like Tesseract are well suited for well-structured, composed documents with simple fonts but struggle with handwriting, stylized, or curved text features.

Feature	Existing Literature	This Study
Focus	OCR accuracy and text recognition improvements	OCR accuracy **and** data security integration
Text Detection Models	EAST, CTPN, YOLOv3, Transformer-based OCR	EAST, CTPN for high-precision text detection
Text Recognition Models	CNN, RNN, CRNN, Transformer-based models	CNN-RNN hybrid model with LSTM and CTC loss
Handwritten Text	Moderate accuracy (typically 85-	Enhanced handwritten text recognition

Support	89%)	
Security Measures	**No built-in security** in most studies	AES hashing ensures text integrity
Data Integrity	No explicit focus on data protection	Encrypted extracted text prevents tampering
Datasets Used	ICDAR 2013, SVT, IIIT 5K, Custom datasets	ICDAR 2013, SVT, IIIT 5K, plus a **custom 5,000-image dataset**
Accuracy on Structured Text	90-95%	**94-96%** (enhanced with preprocessing)
Accuracy on Scene Text	85-92%	**91-94%** (robust against distortions)
Accuracy on Handwritten Text	80-89%	**87-92%** (improved sequence modelling)
Computational Efficiency	Some models are resource-intensive	Optimized for **real-time applications**
Applications	Text detection for **academic, signage, and documents**	**Secure document processing, finance, legal, healthcare**

Table 4: Comparison with Existing Literature

Past research by Wang et al. (2020) and Cao et al. (2020) established the efficacy of deep models like CNNs and RNNs in improving the accuracy of text recognition. However, most previous works concentrate on enhancing extraction performance but neglecting to guarantee data integrity and security issues. This research bridges this loophole through the use of AES encryption to secure that the extracted text cannot be tampered with and is tamper-proof. As compared to transformer-based text recognition models, e.g., of Zheng (2024), which focus on end-to-end text detection and understanding, the current work emphasizes computational efficiency over high accuracy. The experimentation reveals that a hybrid model, combining traditional OCR methods with deep learning and aggressive pre-processing, achieves the best trade-off between performance and resource usage. Furthermore, the work is based on previous work and demonstrates the power of multi-step pre-processing techniques that significantly improve text recognition in real-world heterogeneous environments, including low light, mixed font sizes, and distortions.

5. Conclusion

This study enhances accuracy and security in OCR by deep learning with AES encryption. It preserves text integrity, more so than traditional methods, in finance, healthcare, legal, and government applications. This study developed an advanced Optical Character Recognition (OCR) system incorporating deep learning-text extraction with AES encryption to enhance accuracy and security. With the utilization of CNN-RNN architectures, EAST and CTPN models, and preprocessing, the system efficiently recognized and detected text in diverse conditions such as handwritten and distorted text.

AES hashing integration guaranteed the integrity of extracted text to be tamper-proof. The performance results exhibited better efficiency than conventional OCR mechanisms and had reduced Character Error Rate (CER) and Word Error Rate (WER) with computational efficiency. This work points out the feasibility of integrating artificial intelligence and cryptographic methods for safe text recognition applications. This work successfully addressed major challenges in OCR, including text warping, varying font styles, and security threats in the extracted text. Benchmarking with datasets like ICDAR 2013 and SVT was performed to ensure that the system is better than state-of-the-art OCR systems such as Tesseract in reading complex and handwritten text. The inclusion of AES encryption provides an extra layer of protection, preventing text extracted to be altered undetected. This work makes a new contribution to the field by combining AI-based OCR with cryptographic security, providing a scalable and viable solution for industries looking for secure and precise text extraction.

References

1. Yaqin Wang, "Extraction Algorithm of English Text Information from Color Images Based on Radial Wavelet Transform," IEEE Access, vol. 8, pp. 160050-160064, Aug. 2020, doi: 10.1109/ACCESS.2020.3020621.
2. Tianyu Geng, "Transforming Scene Text Detection and Recognition: A Multi-Scale End-to-End Approach with Transformer Framework," IEEE Access, vol. 12, pp. 40582-40596, Mar. 2024, doi: 10.1109/ACCESS.2024.3375497.
3. Syed Yasser Arafat and Muhammad Javed Iqbal, "Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning," IEEE Access, vol. 8, pp. 96787-96803, May 2020, doi: 10.1109/ACCESS.2020.2994214.
4. Sara Khalid, Jamal Hussain Shah, Muhammad Sharif, Fadl Dahan, Rabia Saleem, and Anum Masood, "A Robust Intelligent System for Text-Based Traffic Signs Detection and Recognition in Challenging Weather Conditions," IEEE Access, vol. 12, pp. 78261-78274, May 2024, doi: 10.1109/ACCESS.2024.3401044.
5. Langcai Cao, Hongwei Li, Rongbiao Xie, and Jinrong Zhu, "A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3," IEEE Access, vol. 8, pp. 176924-176934, Sep. 2020, doi: 10.1109/ACCESS.2020.3025221.
6. Arisa Ueda, Wei Yang, and Komei Sugiura, "Switching Text-Based Image Encoders for Captioning Images With Text," IEEE Access, vol. 11, pp. 55706-55715, June 2023, doi: 10.1109/ACCESS.2023.3282444.
7. Tofik Ali, Mohammad Faridul Haque Siddiqui, Sana Shahab, and Partha Pratim Roy, "GMIF: A Gated Multiscale Input Feature Fusion Scheme for Scene Text Detection," IEEE Access, vol. 10, pp. 93992-94006, September 2022, doi: 10.1109/ACCESS.2022.3203691.
8. Wenyuan Xue, Qingyong Li, and Qiyuan Xue, "Text Detection and Recognition for Images of Medical Laboratory Reports with a Deep Learning Approach," IEEE Access, vol. 8, pp. 407-416, December 2019, doi: 10.1109/ACCESS.2019.2961964.
9. Asghar Ali Chandio, MD. Asikuzzaman, Mark R. Pickering, and Mehwish Leghari, "Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network," IEEE Access, vol. 10, pp. 10062-10078, January 2022, doi: 10.1109/ACCESS.2022.3144844

10. Xiao Qin, Jianhui Jiang, Chang-An Yuan, Shaojie Qiao, and Wei Fan, "Arbitrary Shape Natural Scene Text Detection Method Based on Soft Attention Mechanism and Dilated Convolution," IEEE Access, vol. 8, pp. 122685-122694, July 2020, doi: 10.1109/ACCESS.2020.3007351.
11. Hamam Mokayed et al., "Anomaly Detection in Natural Scene Images Based on Enhanced Fine-Grained Saliency and Fuzzy Logic," IEEE Access, vol. 9, pp. 129102-129109, August 2021, doi: 10.1109/ACCESS.2021.3103279.
12. Bui Hai Phong et al., "A Hybrid Method for Mathematical Expression Detection in Scientific Document Images," IEEE Access, vol. 8, pp. 83663-83684, May 2020, doi: 10.1109/ACCESS.2020.2992067.
13. Phanthakan Kiatphaisansophon et al., "Efficient Text Bounding Box Identification Using Mask R-CNN: Case of Thai Documents," IEEE Access, vol. 12, pp. 49306-49328, April 2024, doi: 10.1109/ACCESS.2024.3383911.
14. Jamshid Bacha et al., "A Deep Learning-Based Framework for Offensive Text Detection in Unstructured Data for Heterogeneous Social Media," IEEE Access, vol. 11, pp. 124484-124498, November 2023, doi: 10.1109/ACCESS.2023.3330081.
15. Randheer Bagi et al., "Cluttered TextSpotter: An End-to-End Trainable Light-Weight Scene Text Spotter for Cluttered Environment," IEEE Access, vol. 8, pp. 111433-111447, June 2020, doi: 10.1109/ACCESS.2020.3002808.