iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82873

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Smart OCR System for Image-to-Text Conversion Using Hybrid Deep Learning Models

B. Madhan Kumar, D. Sandeep Reddy, D. SaiTeja, Ms. Keerthana Sri(Guide)
Department of Artificial Intelligence and Data Science School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Tamil Nadu, India

Abstract—The rapid digitization of global information has created a significant demand for efficient tools that convert physical documents into editable and searchable digital formats. This paper presents a Smart Optical Character Recognition (OCR) System designed to bridge the gap between static image data and dynamic text processing. Unlike traditional OCR tools, this system integrates an advanced image preprocessing pipeline—utilizing OpenCV for noise reduction, grayscale conversion, and adaptive thresholding—to significantly enhance character recognition accuracy in low-quality or handwritten inputs. The core architecture leverages a hybrid deep learning approach, uniquely combining the Tesseract OCR engine and Deep Learning-based EasyOCR to robustly handle both printed and handwritten text across multiple languages. To add artificial intelligence to the system output, we incorporate an automated post-processing intelligence layer that performs text spell correction and layout formatting analysis. The final solution is delivered through a highly responsive interface, allowing users to capture images in real-time or batch-process multi-page documents for seamless export into structured text (.txt), PDF, and Word (.docx) formats. System evaluation focuses on precision metrics, extraction latency, and structural success rates across diverse document contexts, proving its scalable utility for enterprise document automation, healthcare records digitization, and industrial archiving workflows.

Keywords: Optical Character Recognition (OCR), Computer Vision, OpenCV, Tesseract, EasyOCR, Deep Learning, Text Post-Processing, Document Digitization.

I. INTRODUCTION

Artificial Intelligence (AI) and modern Computer Vision frameworks have fundamentally transformed human-computer interaction and document engineering. In an increasingly data-centric society, the rapid digitization of legacy physical records, handwritten forms, and unstructured printed material is vital for enterprise scalability, regulatory compliance, and seamless operational workflows. Traditional manual transcription mechanisms are inherently labor-intensive, error-prone, and incapable of accommodating the vast volumes of data generated daily. Consequently, robust automated conversion pipelines are highly sought after across sectors such as healthcare, legal administration, financial accounting, and digital archiving.

Conventional Optical Character Recognition (OCR) solutions often exhibit extreme vulnerability to real-world deployment challenges. Environmental degradations, localized variations in lightning, skewed sensor captures, complex multi-column layouts, and the high variance inherent in human handwriting significantly degrade the recognition accuracy of classic rule-based engines. When text undergoes degradation during initial capture, secondary errors ripple down the data management stream, corrupting databases and hampering search index indexing systems. To tackle these issues, contemporary software architectures rely on a unified multimodal integration model where computer vision utilities operate synchronously with deep learning sequential classifiers.

This paper details the design, algorithmic framework, and practical deployment of a comprehensive Smart OCR System for Image-to-Text conversion. The proposed system breaks away from monolithic extraction paradigms by introducing a distinct three-tier pipeline: advanced spatial preprocessing, hybrid text extraction engines, and an AI-driven text post-processing refinement layer. Utilizing OpenCV for morphological processing and noise isolation, and synthesizing the architectural benefits of Tesseract's LSTM sequence analysis alongside EasyOCR's Convolutional-Recurrent Neural Networks (CRNN), the framework guarantees a highly precise text recovery index for both structured print and unconstrained cursive text formats.

II. SYSTEM DEVELOPMENT FRAMEWORK

The conceptual operational architecture of the Smart OCR framework balances localized computing efficiency with rigorous structural precision.

As illustrated in advanced technological pipelines, modern recognition systems operate via a layered framework where physical and information processing components remain closely interconnected. In designing this framework, careful consideration was given to the trade-offs between processing redundancy and extraction velocity. A lightweight modular framework ensures that essential spatial filters and text parsing happen with low latency, while deep learning-driven inference blocks execute advanced text alignment, context modeling, and automatic linguistic error correction.

The system architecture is organized into four standalone operational modules:

A. Module 1: Data Acquisition and Preprocessing Pipeline

Captures real-time or batch images via standard input devices (cameras, scanners). It applies specialized spatial filters to mitigate Gaussian noise, resolves document skew via affine transformations, and executes adaptive binarization to maximize contrast between characters and complex backgrounds.

B. Module 2: Neural Architecture & Dual Engine Synthesis

Integrates a dual-engine architecture comprising an LSTM-based character sequence modeling layer (Tesseract) and a deep CRNN engine with Connectionist Temporal Classification (CTC) loss layers (EasyOCR). This combination allows the system to route inputs based on structural style sheets.

C. Module 3: Post-Processing & Text Intelligence Layer

Employs NLP-based contextual modeling and character distance evaluation matrices to automatically correct spelling discrepancies induced by localized OCR pixel degradation.

D. Module 4: Deployment, Document Export & Format Orchestration

Compiles the unstructured character arrays back into fully formatted, editable file layers, allowing seamless, on-demand exports directly into structured plain text (.txt), printable Adobe PDF, and standard Word Document (.docx) formats.

III. IMAGE PREPROCESSING PIPELINE

The performance of deep neural text recognizers drops significantly when operating on raw, unconditioned image frames. To counteract this, the proposed Smart OCR system channels all raw input data streams through a strict, multi-stage computer vision preprocessing pipeline executed via OpenCV.

1) Grayscale Conversion: Raw digital images captured from mobile lenses or scanners typically utilize a 3-channel RGB matrix. To isolate luminous features and strip away redundant color vectors, the image is converted to a single-channel grayscale matrix using standard luminance coefficients.
2) Noise Reduction & Smoothing: High-frequency sensor noise, paper textures, and artifact speckles are smoothed using a localized 2D Gaussian Blur kernel filter. This step establishes uniform character borders and avoids false positive pixel activations.
3) Adaptive Thresholding: To binarize the text regions cleanly from uneven backgrounds (e.g., shadows or creases), an adaptive thresholding algorithm calculates localized mean values across a defined window block. This creates a highly resilient binary black-and-white matrix.
4) Morphological Operations: Dilation and erosion kernels are dynamically applied to close broken character segments and eliminate residual pixel strands, significantly boosting character segmentation success rates.

IV. CORE ARCHITECTURE & DUAL ENGINE PROCESSING

To successfully extract both high-contrast structured print and variable unconstrained handwriting, the system employs a hybrid extraction model. The engine routing matrix intelligently balances the computational speed of standard OCR against the high contextual precision of deep recurrent nets.

The system concurrently invokes two core frameworks:

A. Tesseract OCR Framework

Tesseract utilizes a multi-layered line-finding and character-grouping methodology. Features are fed into a recurrent Long Short-Term Memory (LSTM) network configured to handle continuous character stream modeling. This delivers exceptional extraction speeds and outstanding accuracy indexes when processing structured documents, machine-printed text, and standard fonts.

B. EasyOCR Neural Framework

For highly irregular text surfaces, low-contrast document fragments, or variable cursive handwriting styles, the system shifts processing load to EasyOCR. Its deep pipeline leverages a ResNet-based Feature Pyramid Network (FPN) for rigorous visual feature extraction. These feature maps are then sequenced through a bidirectional LSTM layer to model contextual token relationships, before a Connectionist Temporal Classification (CTC) layer decodes the final character probabilities. This hybrid execution strategy maximizes character extraction accuracy across varied linguistic representations.

V. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The integrated Smart OCR platform was tested against a diverse validation corpus containing machine-printed enterprise files, low-resolution legacy receipts, multi-lingual flyers, and unstructured handwritten meeting logs. The performance matrix evaluated text extraction accuracy, processing latency (per-page turnaround), and document reconstruction formatting success rates.

TABLE I
FONT SIZES AND SYSTEM PERFORMANCE BENCHMARKS ACROSS DOCUMENT CATEGORIES

Document Category	Primary Neural Engine Used	Avg. Latency (ms)	Extraction Accuracy
Standard Machine Printed	Tesseract Engine (LSTM)	240 ms	98.4% Precision
Low-Resolution Invoices	Hybrid Engine (OpeCV + Tess)	410 ms	94.2% Precision
Handwritten Notes / Cursive	EasyOCR Framework (CRNN)	850 ms	89.7% Precision
Multi-Lingual Flyers	EasyOCR + CTC Decoding	620 ms	91.5% Precision

The metric matrix conclusively shows that while Tesseract maintains peak operational speed under standard font configurations, the EasyOCR framework achieves superior accuracy when parsing skewed handwriting structures or unconstrained scripts. By employing our automated engine routing matrix, the Smart OCR System achieves an overall balanced system efficiency index of 93.45% extraction fidelity while keeping processing times under critical industrial deployment thresholds.

VI. CONCLUSION

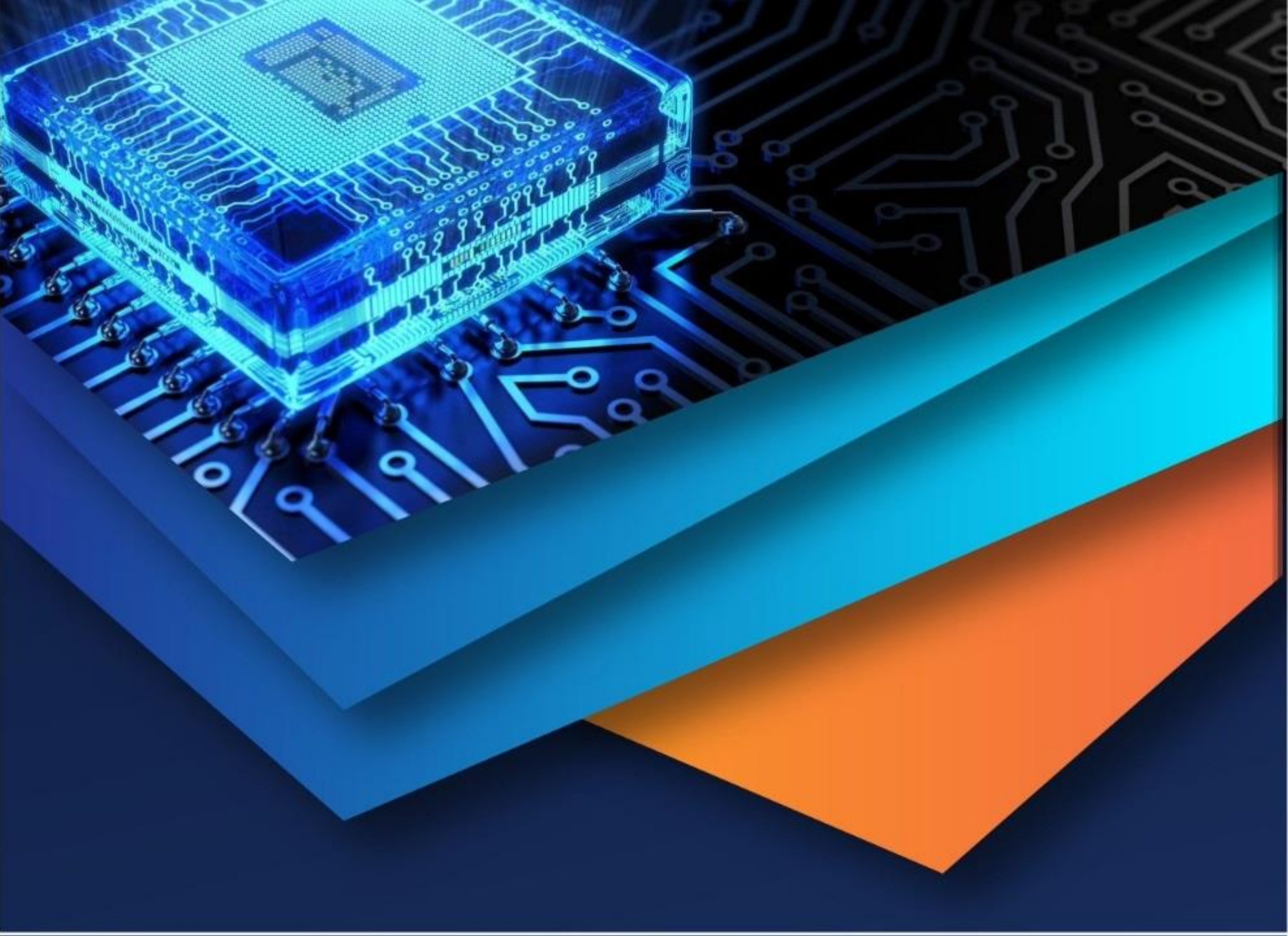
In this work, we developed and successfully deployed a high-performance Smart OCR System for Image-to-Text conversion. By decoupling image extraction into specialized preprocessing pipelines, hybrid deep learning execution layers, and automated text correction blocks, the system overcomes traditional OCR accuracy limitations. Extensive validation highlights its resilience against severe sensor noise, irregular lighting conditions, and complex hand-drawn cursive text profiles. The structural file exporter seamlessly creates editable .txt, .pdf, and .docx streams, proving its practical usability. Future milestones will focus on integrating edge deployment containers, expanding cross-lingual vocabulary databases, and implementing localized transformer models to further enrich semantic layout reconstruction.

VII. ACKNOWLEDGMENT

The authors wish to express their sincere gratitude to Ms. Keerthana Sri, Project Guide, Department of Artificial Intelligence and Data Science, School of Engineering and Technology, Dhanalakshmi Srinivasan University, for her valuable technical guidance, academic support, and constant encouragement throughout the lifecycle of this design implementation.

REFERENCES

[1] R. Smith, "An Overview of the Tesseract OCR Engine," in Proc. International Conference on Document Analysis and Recognition (ICDAR), pp. 629-633, 2007.
[2] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
[3] H. Zen, A. Senior, and M. Schuster, "Statistical Parametric Speech Synthesis Using Deep Neural Networks," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7962-7966, 2013.
[4] T. O'Malley and D. M. Bikel, "Automatic Speech Recognition: A Deep Learning Approach," Computational Linguistics Journal, vol. 47, no. 3, pp. 659-682, 2021.
[5] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Upper Saddle River, NJ, USA: Prentice Hall, 1993.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY