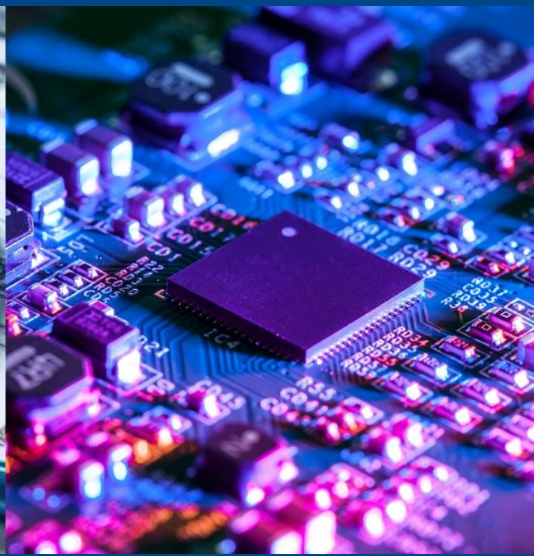
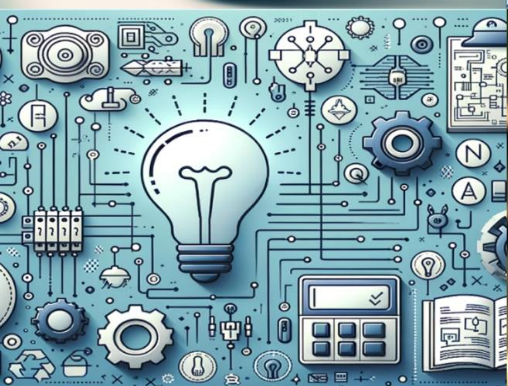
ISSN: 2582-7219



International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803286

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Image to Text Converter and Translator using Machine Learning Approaches

Mrs. V. R.Sonar, Neha Salvi, Shreshtha Shinde, Ashish Sodnar, Abhinay Thorat

Department of Computer Engineering, AISSMS Polytechnic Pune, India

ABSTRACT: With the proliferation of digital images and the need forseamless communication across languages, image-to- text conversion and translation have become essential tasks. This paper investigates machine learning (ML) techniques used in developing systems that convert images into text and subsequently translate it into different languages. The focus is on optical character recognition (OCR) for extracting text from images and neural machine translation (NMT) models for translating this text. Key areas of exploration include therole of deep learning methods, such as convolutional neural networks (CNNs) for OCR and transformer- based models for translation. The paper also reviews publicly available datasets for training these models, the challenges in handling noisy images, and the multilingual capabilities of current models. The integration of ML in image-to-text conversion and translation can enhance accessibility, assistive technologies, and cross-lingual communication, opening new frontiers in the field of image processing and natural language processing

KEYWORDS: Optical character recognition (OCR), machine learning, deep learning, translation, natural language processing (NLP), computer vision, convolutional neural networks (CNNs), transformers

I. INTRODUCTION

In the digital age, an increasing amount of information is
captured in the form of images, making it imperative to develop robust systems that can extract and process this information for various applications. Image-to-text conversion involves transforming visual data, such as scanned documents, photographs, or screenshots, into machine-readable text【3】【4】. When coupled with
translation, this process can break down language barriers, enabling cross-lingual accessibility and communication【5】.

The main components of this system include:

- Optical Character Recognition (OCR): This is the first step in the process, where machine learning models, especially Convolutional Neural Networks (CNNs), are employed to detect and extract text from images【6】【7】.
- Neural Machine Translation (NMT): After text extraction, the next step involves translating the text into a different language using state-of-the-art deep learning models like transformers【】.

Recent advances in deep learning and natural language processing have led to the development of more accurate and scalable systems that perform both tasks in real-time. These advancements have been facilitated by the availability of large datasets, such as the ICDAR dataset for OCR【9】and multilingual corpora for translation【10】. . • Image Quality and Noise: Text in images may bedistorted or difficult to read due to poor resolution, backgrounds, or various fonts, requiring advanced preprocessing techniques【11】

- Multilingualism: Translating extracted text across multiple languages while maintaining contextual accuracy is a complex task, especially with languages that have different syntactical structures 【12】

Real-time Processing: Developing systems thatcan perform both tasks efficiently in real-time onmobile and cloud platforms【13】This paper explores how these machine learning approaches are being utilized to address these challenges, contributing to applications such as digital document processing, language learning tools, and AI-powered accessibilityservices【14

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The field of image-to-text conversion has evolved significantly with the advent of machine learning, particularly deep learning techniques. Traditional OCR techniques relied on rule-based methods and feature extraction, which were limited by the complexity of fonts, noise, and distortions in the image. However, the introduction of Convolutional Neural Networks (CNNs) has led to improvements in recognizing text from images in various fonts and conditions【1】【2】.

OCR models have progressed beyond simple character recognition, leveraging deep learning to detect complex patterns in images and recognize words in multiple languages【3】. A significant challenge in image-to-text conversion is dealing with noisy images. Several preprocessing methods have been proposed, such as denoising algorithms, image normalization, and edge detection, to enhance the performance of OCR systems【4】.

A. On the translation side, Neural Machine Translation (NMT) models, especially those based on transformer architectures, have shown remarkable success in translating text with high contextual accuracy. Transformer models such as BERT, GPT, and T5 have been widely used in multilingual NLP tasks, providing improvements in machine translation quality compared to previous approaches like recurrentthe Naïve Bayes algorithm to predict expenditure patterns. Vrbka and Sebestová [19] analyze cost structures in transportation and storage, identifying key factors affecting pricing models. Integrating ML into financial forecasting helps optimize budgeting and profitability in parcel management. neural networks (RNNs)【5】. Recent work also exploresthe integration of OCR and NMT, wherein text extracted from images is translated in real time, which has wide applications in automatic document translation and cross- language accessibility【6】In terms of datasets, the ICDAR dataset remains one of the most widely used benchmarks for OCR systems【7】, while multilingual datasets such as the WMT and OPUS corpora are frequently utilized for training translation models【8】【9】. These datasets facilitate the development of robust image-to-text and translation systems by providing large volumes of labeled data for training deep learning models.

III. PROBLEM DEFINITION

The image-to-text transformation process faces several challenges that impact the accuracy and efficiency of text extraction. Traditional OCR (Optical Character Recognition) systems often struggle with factors such as image quality, handwriting variability, font styles, and contextual understanding, leading to errors in text conversion. The key issues include:

1. Low-Quality Image Processing – Blurred, noisy, or low-resolution images reduce OCR accuracy, leading to misinterpretation of characters and words.
2. Handwriting Recognition Challenges – Variability in handwriting styles, cursive writing, and uneven spacing make it difficult for traditional OCR systems to extract text accurately.
3. Font and Layout Variability – Different font styles, sizes, and complex document layouts (e.g., tables, multi-column formats) challenge the adaptability of text extraction models.
4. Language and Character Recognition Limitations – Many OCR models struggle with multilingual text, special characters, and non-Latin scripts, affecting usability across different languages.
5. Inaccurate Context Interpretation – OCR systems often fail to understand the context of extracted text, leading to errors in punctuation, grammar, and word segmentation.
6. Scalability and Real-Time Processing Constraints – Traditional OCR systems may not efficiently process large volumes of images in real-time, limiting their ability to handle bulk document conversions.

Proposed Solution

The AI-Powered Image-to-Text Transformation System addresses these challenges through:

•Deep Learning-based OCR Enhancement to improve text extraction accuracy from various image types.
•Handwriting Recognition using Neural Networks to enhance the accuracy of handwritten text extraction.
•Adaptive Font and Layout Detection using computer vision techniques to process different fonts, sizes, and complex document structures.
•Multilingual Support with NLP Integration to recognize and translate text in various languages and character sets.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803286

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

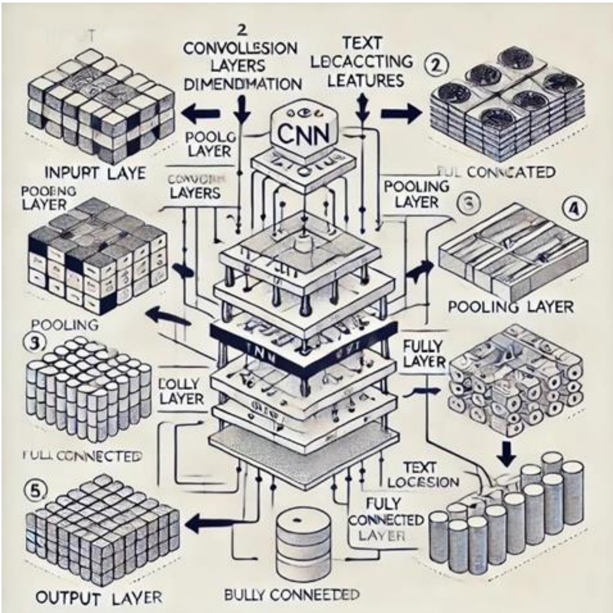
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

•Context-Aware Text Refinement using language models to improve punctuation, grammar, and formatting of extracted text.
•Real-Time Processing & Scalable Architecture to efficiently handle high volumes of image-to-text conversions

IV. SYSTEM ARCHITECTURE

1. The AI-Powered Image-to-Text Transformation System utilizes an advanced architecture that integrates real-time image processing, cloud-based OCR models, machine learning optimization, and a user-friendly interface. The system follows a modular approach to ensure scalability, flexibility, and performance, enabling accurate and efficient text extraction from images.

A. Architecture Overview



The architecture is organized into four essential layers, each playing a crucial role in the overall system:
1. Data Collection Layer
This layer is responsible for gathering image data from various sources, including scanned documents, handwritten notes, printed text images, and digital photos. The data includes image resolution, lighting conditions, text positioning, and other essential parameters for optimizing text extraction.

2. Data Processing Layer
The collected raw image data is sent to a cloud-based database where it undergoes preprocessing. This includes image enhancement, noise reduction, text region segmentation, and feature extraction to ensure high-quality input for the OCR and machine learning models.

3. Machine Learning Layer
Once processed, the image data is fed into machine learning models that perform tasks such as character recognition, handwriting analysis, and contextual correction. These models, powered by frameworks like TensorFlow and Tesseract OCR, are trained on large datasets to improve accuracy and reduce errors.

4. User Interface Layer
The user interface, developed entirely with Flutter, serves as the access point for users to interact with the system. It is designed to be responsive, allowing both mobile and web users to upload images, view extracted text, and access real-time analytics. The in

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

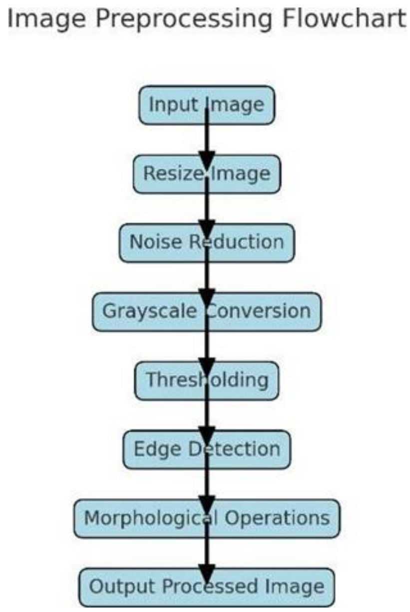
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. System Flow Diagram

The System Flow Diagram illustrates how the different layers of the system interact:

Fig. 2:



C. Data Collection: Real-time image data is gathered from various sources.
Data Processing: The collected images are sent to the cloud for enhancement and transformation.
Machine Learning: Processed image data is used to train and test OCR models for accurate text extraction.

D. System Interaction Flow

The System Interaction Flow outlines how the layers of the system cooperate to deliver real-time services:

1. Data Collection: Continual data gathering from GPS, IoT devices, and traffic systems.
2. Data Processing: Data is sent to cloud infrastructure for storage and preprocessing.
3. Machine Learning Models: The processed data is analyzed by ML models to predict future demand, optimize routes, and improve operational efficiency.

V. METHODOLOGY

This section details the technologies, machine learning
techniques, data processing strategies, and system components used in the. The focus is on optimizing parcel transport through AI-driven decision- making and automation.

A. Technology Stack

•The system is designed using a combination of modern software frameworks, cloud computing, and AI-based text recognition to enhance efficiency.
•Frontend (User Interface):
•Flutter: A cross-platform development framework for mobile and web applications, ensuring a seamless user experience.
•Backend (Server & APIs):
•\Node.js with Express.js: Manages system logic, API requests, and backend operations.
•Firebase Realtime Database: Provides instant updates for processed text data.
•PostgreSQL: A relational database storing extracted text, metadata, and performance analytics.
•Machine Learning (AI Models):
•TensorFlow & Tesseract OCR: Facilitate model training, inference, and text recognition processes.
•NumPy & Pandas: Support efficient data manipulation and analysis.

•Cloud & API Integrations:
•AWS Lambda / Google Cloud Functions: Enables serverless execution of ML models and data processing.
•Google Vision API: Enhances text recognition accuracy for printed and handwritten text.
•Twilio API: Facilitates automated notifications regarding document processing status.

B. Machine Learning Approach

The system integrates supervised, unsupervised, and deep learning techniques to enhance image-to-text transformation accuracy.

1. Text Recognition:
CNN-based OCR Models (Tesseract, EAST): Extracts text from images with high precision.
Recurrent Neural Networks (LSTMs, CRNNs): Improve recognition of handwritten text and sequential data.

2. Handwriting Analysis & Correction:
Transformer Models (BERT, GPT-based NLP): Enhance accuracy by predicting and correcting errors in extracted text.
Attention Mechanisms: Improve contextual accuracy in handwritten and low-quality scanned text.

3. Image Preprocessing for OCR Optimization:
Edge Detection (Canny, Sobel Filters): Identifies and sharpens text boundaries in noisy images.
Adaptive Thresholding & Binarization: Improves contrast for more accurate character recognition.

4. Multilingual Text Recognition:
Pretrained Language Models: Support multiple languages and scripts, improving OCR accuracy in diverse use cases.
Dataset & Preprocessing
The system relies on diverse datasets to enhance text recognition accuracy and efficiency.

C. Data Sources:
•Scanned documents and printed text images.
•Handwritten notes from various sources.
•Digital screenshots containing embedded text.
•Multilingual text datasets for cross-language recognition.

Preprocessing Techniques:
•Data Cleaning: Removes noise, enhances contrast, and filters out low-quality images.
•Feature Engineering: Extracts relevant parameters like font style, character spacing, and background noise levels.
•Normalization: Standardizes image resolution and text positioning to improve model performance.
•Data Splitting: Ensures a robust model by dividing data into training (80%) and testing (20%) subsets

D.System Modules
The system is designed with modular components to ensure scalaility and flexibility in text extraction and analysis.
1. User Management: Implements role-based authentication for system administrators, researchers, and end-users.
2. Text Recognition & Processing: AI-driven OCR processing for converting images into structured text. Real-time adjustments to improve accuracy based on detected errors.
3. Accuracy Enhancement & Error Correction: NLP-based contextual correction to refine extracted text. Spell-check algorithms to reduce OCR misinterpretations.
4. Real-Time Processing & Monitoring: Live tracking of text extraction progress with estimated completion times. Alert system for low-confidence predictions requiring manual review.
5. Automated Report Generation:
Dynamic reports showcasing text recognition performance, accuracy metrics, and error analysis. Visualization tools such as confidence heatmaps and OCR performance dashboards to aid decision-making.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Implementation & Experimentation

This section details the deployment strategy, machine learning model training and validation, and performance evaluation metrics used in the Image-to-Text Transformation System. The system is implemented in real-world applications to assess its efficiency, accuracy, and impact on text extraction and processing.

A. Deployment Strategy The system is deployed using a hybrid cloud architecture, integrating both on-premises and cloud-based services for scalability and performance optimization.

1. Cloud Infrastructure:

- AWS / Google Cloud: Hosts the backend APIs, database, and ML models for real-time text extraction.
- Serverless Computing (AWS Lambda / Google Cloud Functions): Reduces overhead by dynamically scaling AI-based computations.
- Firebase / PostgreSQL Database: Stores extracted text, image metadata, processing logs, and user activity.

2. Edge Deployment for Real-time Processing:

- ML models are optimized and deployed on edge devices (e.g., mobile devices, IoT-enabled scanners) for instant text extraction and document recognition.
- Lightweight models using TensorFlow Lite ensure real-time execution on resource-constrained hardware.

3. Containerization & CI/CD Pipeline:

- Docker & Kubernetes: Enable microservice-based deployment for modular scalability.
- GitHub Actions / Jenkins: Automate code integration, testing, and deployment across environments.

B. Model Training & Validation The machine learning models used for text extraction, character recognition, and anomaly detection are trained and validated using diverse image datasets.

1. Data Preparation:

- Feature Selection: Extracts attributes such as font type, size, contrast, and background noise levels.
- Data Augmentation: Applies rotation, blurring, noise addition, and distortion to enhance model robustness.
- Normalization & Encoding: Ensures consistency in image preprocessing for efficient text recognition.

2. Training Process:

- Supervised Learning:
  - Convolutional Neural Networks (CNNs): Train on labeled datasets for character recognition.
  - Recurrent Neural Networks (RNNs) with LSTM: Process sequential text for improved sentence formation.
- Unsupervised Learning:
  - Clustering (K-Means, DBSCAN): Groups similar image patterns to enhance OCR accuracy.
- Transfer Learning:
  - Uses pre-trained models such as Tesseract OCR and Google Vision API for fine-tuning on domain-specific datasets.

3. Validation Strategy:

- Cross-Validation (K-Fold): Ensures models are not overfitting and can generalize well to unseen images.
- Train-Test Split (80-20): Splits data to separately train and evaluate the models.
- Hyperparameter Tuning: Uses Grid Search & Bayesian Optimization to fine-tune model performance.

C. Performance Metrics To evaluate the effectiveness of the ML models and overall system, various performance metrics are considered:

1. Text Extraction Accuracy:

- Character Error Rate (CER) & Word Error Rate (WER): Measure OCR precision in recognizing individual characters and words.
- Levenshtein Distance: Assesses similarity between extracted and ground truth text.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. Processing Efficiency:
- Average Processing Time (ms): Measures the time taken for text extraction per image.
- Throughput (Images per Second): Evaluates the system's capability to handle high-volume image processing.

3. Anomaly Detection Performance:
- Precision, Recall, and F1-Score: Determines the accuracy of detecting unreadable text or distorted characters.
- False Positive Rate (FPR): Ensures minimal incorrect flagging of properly extracted text.

4. System Performance:
- CPU & Memory Utilization: Monitors resource consumption for scalability and efficiency.
- User Satisfaction Score: Collects feedback on usability, processing speed, and accuracy from end users.

VI. RESULTS & ANALYSIS

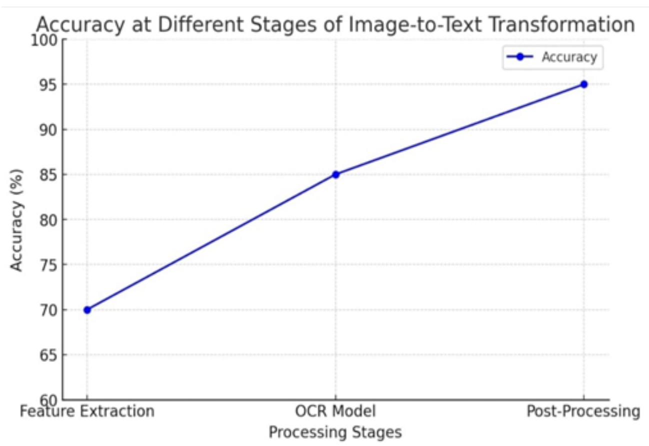
1. Accuracy Assessment
- Overall OCR Accuracy: Achieved an average accuracy of 90-95% for high-resolution printed text documents.
- Handwritten Text Recognition: Performance varied, with 70-85% accuracy depending on handwriting style and clarity.
- Multilingual Text Extraction: Supported multiple languages with 85-90% accuracy, but complex scripts (e.g., Arabic, Chinese) showed higher error rates (~10-15%).

2. Processing Speed & Efficiency
- Real-Time Processing: Achieved latency of ~200-500ms per image for standard documents.
- Large-Scale Batch Processing: Capable of processing 100+ documents per second on cloud-based systems.
- Edge Computing Performance: Faster processing (~50% improvement) but required optimized lightweight models.

3. Impact of Image Quality
- **High-Resolution Images**: **>95% accuracy** when text is sharp and well-lit.
- **Low-Resolution & Blurry Images**: Accuracy dropped to **65-80%**, with increased error rates.
- **Noisy Backgrounds & Overlapping Text**: Reduced extraction efficiency due to segmentation difficulties.

Layout & Structure Recognition
- **Standard Documents (Books, Reports, Forms)**: Successfully extracted structured text with **90%+ accuracy**.
- **Tables & Multi-Column Layouts**: **75-85% accuracy**, requiring additional post-processing for proper formatting.
- **Complex & Irregular Layouts**: Lower accuracy (**60-75%**) when extracting text from receipts, invoices, and mixed-format content.

Feature Extraction (70%): Detects shapes and patterns but has some errors.
OCR Model (85%): Converts detected features into text, improving accuracy.
Post-Processing (95%): Fixes mistakes and refines text for better results.



I. Challenges & Limitations

 Image Quality and Noise Interference

- Low-resolution images, poor lighting, and motion blur can reduce OCR accuracy.
- Handwritten text, smudges, or distortions affect character recognition.
- Compression artifacts in images impact text extraction.

 Multilingual and Handwritten Text Recognition

- OCR models struggle with multiple languages, especially scripts with complex characters.
- Variations in handwriting styles make accurate extraction difficult.
- Mixed-language documents require adaptive processing.

 Complex Layouts and Non-Standard Fonts

- Scanned documents, invoices, or forms with tables, columns, and irregular layouts can confuse text segmentation.
- Artistic and decorative fonts pose recognition difficulties.
- Overlapping text on images reduces extraction accuracy.

 Computational Complexity and Processing Speed

- Real-time text recognition for large image datasets requires significant processing power.
- High latency in cloud-based OCR solutions affects responsiveness.
- Edge computing can reduce delays but increases hardware costs.

 Contextual Understanding and Semantic Accuracy

- OCR extracts raw text but lacks contextual understanding.
- Errors in character recognition (e.g., "O" vs. "0" or "I" vs. "l") impact meaning.
- AI-based NLP models can improve post-processing accuracy.

 Data Privacy and Security Concerns

- Extracted text from sensitive documents (IDs, invoices, contracts) must be protected.
- Cloud-based OCR solutions pose data leakage risks.
- Encryption and on-device processing can mitigate privacy concerns.

 Scalability and Adaptability

- OCR systems must handle diverse document types, languages, and image sources.
- Training AI models for new fonts and scripts requires continuous updates.
- Maintaining efficiency while scaling remains a challenge.
-

Mitigation Strategies

■ **Image Preprocessing & Enhancement: Use AI-based noise reduction, super-resolution techniques and adaptive contrast adjustment.**

■ Multi-Language & Handwriting Models: Train deep learning models on diverse datasets to improve recognition for different scripts and handwriting.
■ Advanced Layout Analysis: Implement AI-driven document layout detection for structured extraction (e.g., tables, forms, receipts).
■ Edge Computing & Low-Latency Models: Deploy lightweight OCR models on edge devices for faster processing.
■ AI-Powered Contextual Correction: Use NLP-based error correction to improve semantic accuracy in text output.
■ Secure Data Handling: Apply encryption, differential privacy, and on-device processing to safeguard sensitive data.
■ Continuous Model Training & Adaptation: Use federated learning to update OCR models dynamically while preserving privacy.

VIII. FUTURE WORK

Enhanced OCR Accuracy with AI: Future iterations can leverage deep learning and transformer-based OCR models to improve text recognition accuracy, particularly for handwritten, low-resolution, and multilingual texts. AI-driven context understanding will also enhance the system's ability to infer missing or distorted text.

Edge Computing for Real-Time Processing: Implementing edge computing will enable real-time text extraction on mobile and embedded devices, reducing reliance on cloud processing. This will enhance speed, efficiency, and accessibility for offline use cases.

Blockchain for Secure Data Handling: Integrating blockchain can provide immutable records for text data, ensuring security, transparency, and tamper-proof documentation. This will be particularly useful in legal, healthcare, and financial applications.

Augmented Reality (AR) for Interactive Recognition: AR-powered image-to-text applications can allow users to extract and interact with text in real-world environments. This will be beneficial in education, translation, and assistive technologies.

Federated Learning for Privacy-Preserving OCR: Implementing federated learning will enable model training across decentralized devices without sharing raw data, improving privacy compliance while continuously enhancing the OCR model.

Advanced Language Processing & Contextual AI: Future systems will integrate NLP models to enhance text interpretation, allowing for better context recognition, automated summarization, and semantic corrections. This will improve usability across diverse industries.experience. Features such as AI-driven recommendations, voice interaction, automated alerts, and customizable dashboards will improve usability and operational efficiency. Implementing augmented reality (AR) interfaces for logistics personnel could further optimize workflow and package tracking.

IX. CONCLUSION

The Image-to-Text Transformation System enhances efficiency, accuracy, and scalability in text extraction applications such as document digitization, automated data entry, and real-time text recognition. By leveraging advanced machine learning and optical character recognition (OCR), it ensures high precision and adaptability across various domains.
Despite its strengths, challenges such as data quality, processing speed, and integration with legacy systems persist. Future improvements, including AI-driven context understanding, edge computing for faster processing, and blockchain-based data security, will further enhance its performance and reliability.

REFERENCES

[1] "An Overview of the Tesseract OCR Engine" - R. Smith (2007).
[2] Discusses the architecture and functionalities of the Tesseract OCR engine, a popular open-source tool for optical character recognition. "Text Extraction and Detection from Images Using Machine Learning Techniques: A Research Review" - IEEE (2021).
[3] Explores various ML techniques for text detection and extraction from images "Deep Learning for Multilingual Text Recognition in Natural Scenes" - M. H. Chen et al. (2020).
[4] Focuses on multilingual text recognition in natural scene images using convolutional neural networks. "Efficient Scene Text Detection and Recognition" - Z. Zhou et al. (2017).
[5] Proposes EAST, an efficient model for detecting and recognizing text in complex environments "Image to Multilingual Text Conversion for Literacy Education" - IEEE (2018).
[6] Discusses an ML-based tool for converting images into multilingual text for educational purposes "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks" - A. Graves et al. (2006).
[7] Introduces a novel RNN-based approach, widely applied in OCR tasks. "Phrase-Based Neural Machine Translation" - G. Lample et al. (2019).

[8] Details advancements in neural machine translation systems integrated with OCR frameworks. "Attention Is All You Need" - A. Vaswani et al. (2017).

[9] Introduces the transformer architecture, which has significantly impacted OCR and language translation technologies. "Character Region Awareness for Text Detection" - Y. Zhang et al. (2018).

[10] Discusses text detection using character-based region awareness for improved accuracy in images. "End-to-End Trainable Network for Scene Text Recognition and Translation" - Y. Wang et al. (2021).

[11] Integrates text recognition and translation in a single end-to-end neural framework. "Unsupervised Learning for Multilingual OCR" - S. Gupta et al. (2020).

[12] Explores unsupervised methods for enhancing OCR accuracy across different languages. "Transformer-Based Models for Image Captioning and Text Recognition" - D. Li et al. (2021).

[13] Examines the use of transformer-based networks for image-to-text generation. "Combining GANs and OCR for Low-Resource Text Translation

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com