

E-13, UPSIDC Site-IV, Kasna Road, Greater Noida - 201308, India

Phone: +91.120.4296878 | Website: [www.khuranaandkhurana.com](http://www.khuranaandkhurana.com/) [www.iiprd.com](http://www.iiprd.com/) Email: [info@khuranaandkhuraana.com](mailto:info@khuranaandkhuraana.com) | [iiprd@iiprd.com](mailto:iiprd@iiprd.com)

Indian Offices: Noida | New Delhi | Bangalore | Mumbai | Pune | Mumbai | Hyderabad | Jalandhar | Chennai International Offices: US |Bangladesh | Sri Lanka | Malaysia | Vietnam | Myanmar | Nepal | Malaysia

# INVENTIVE DISCLOSURE - CONFIDENTIAL

1. **Proposed Title of the Invention** (Not more than 10 to15 words)

**Text**

**AI-Enhanced OCR System for Accurate Text Recognition and Processing**

1. **Proposed Abstract of the Invention** (Kindly explain the crux of the invention in about150 to 200 words.)

This system advanced technology identifies ways to enhance Tesseract OCR's ability to transform scanned PDF documents into textual data. Our system uses LLM text correction to resolve typical OCR errors while handling both character recognition mistakes and poor document formatting. First PDF page images are created and passed to Tesseract OCR which recovers text data. After text generation LLMs polish the output to deliver finished content of high quality and easy readability.

Alongside better text accuracy this system lets visually impaired users listen to text content through its text-to-speech technology. The system reads corrected text aloud to let blind and visually impaired users access written information. The system helps everyone access important details by presenting information in ways print versions often block.

The system uses multiple processing cycles without delay to make document handling faster and more reliable. This technology benefits numerous activities such as academic research while making legal and personal content easier to use. The invention joins superior OCR technology with automated text editing and accessibility functions to enhance document processing systems across multiple fields. From this update users gain smarter document searching benefits and more comfortable user interactions.

# Key Words:

Coherent Text, Visually Impaired ,Asynchronous Processing, Structured Outputs, Token Management, Markdown Formatting, Quality Assessment, Model Integration, Workflow Optimization.

# Background of the Invention:

**What are the present technologies that exist in the field of your invention and what are the limitations of the same? (Present state of Art)**

Traditional OCR tools like Tesseract represent most current imaging technology to transform document scans into text that computers can read. The popularity of Tesseract comes from its open-source design and performance but it shows real issues identifying text in complicated documents and fixing OCR mistakes. OCR-generated text flaws such as character recognition faults and formatting errors alter the extracted text in ways that reduce its value for users.

When OCR tools produce inaccurate results manual human editors need to fix them yet this process takes long and requires a lot of work. The human effort required for these manual procedures makes them impractical for handling excessive document quantities because they slow down processing and allow errors to go undetected. OCR systems today do not include built-in features for helping visually impaired users access content which restricts their interaction with the information.

To help visually impaired users text-to-speech systems exist alongside OCR systems but operate separately from each other. The systems stop working well together when users need to switch between them to find information. Our LLM-Aided OCR Project merges Large Language Models with OCR systems to upgrade output quality and accuracy while enabling speech-to-text functionality for better accessibility by visually impaired users. Our new system helps overcome technology restrictions while giving full document handling services to all users.

**Present State of Art**

**Traditional OCR Systems:**

OCR systems process scanned documents to create text files.

OCR system performance decreases when dealing with intricate document formatting while its text reading accuracy remains substandard.

**Manual Proofreading:**

Manually verifying OCR errors takes too much work to process large sets of documents.

The process takes longer than needed and does not catch all mistakes.

**Text-to-Speech Technologies:**

Text-to-Speech features help sight impaired users while they cannot work directly with OCR software.

Users must move back and forth between multiple applications which makes navigating the system harder.

**Accessibility Features:**

OCR technology today lacks built-in functions that help visually impaired users navigate scanned documents.

Users who depend on these tools face barriers when trying to work with online content.

**Error Correction Mechanisms:**

OCR mistakes in documents need manual review because adjustment mechanisms are absent in present systems.Users will face less accurate results and have reduced usage of their data when errors occur in OCR extraction.The LLM-Aided OCR Project modifies its system to make OCR better through advanced language models and then adds automatic mistake fixing tools plus accessibility functions for blind users.

Latest OCR Technology Tools Help Blind People Better Understand Scanned Text

Large Language Models help Optical Character Recognition better recognize text from scanned documents through advanced technology updates in this field. The modern OCR technology includes Tesseract which turns scanned text images into machine-readable format. These systems work better thanks to LLMs that help OCR explain text errors within their full context. Text-to-speech programs turn the adjusted text into spoken words to help blind people use the system more easily. New developments in text conversion now combine cloud services with machine learning models to detect text more accurately. These services work fast on many documents and fix mistakes using LLMs in real time. By scanning documents anywhere through mobile apps with OCR and TTS technology users can obtain information wherever they need it.

These existing systems do not work as well as needed

Our text recognition tools suffer from basic performance problems when used to help visually impaired users. OCR technology still produces too many errors when working with text from difficult images and hard-to-read handwriting. The issues with weak context detection remain unresolved by LLMs because they cannot handle complex situations well.Connecting LLMs with OCR tools works well but not perfectly. Most current systems need human help to fix errors yet this helps process takes too long for people who need fast access to data. LLMs show limited consistency in performance by generating weak results for texts that use technical or industry-specific language.TTS tools improve accessibility yet their output may feel unnatural and hard to use. User problems arise when speech synthesis systems provide poor sound quality and challenging document navigation during long texts containing complex wording. The system's difficulties hamper the visual impaired user's complete engagement with the content. Our LLM-Aided OCR Project creates a better and easier way for people to process documents and access content despite these system difficulties.

# What problems does the invention address and how your Invention is able to overcome the limitations/ problems of the existing technologies?

# The innovation helps identify these target areas

# Through the LLM-Aided OCR Project users gain better benefits compared to current Optical Character Recognition tools. Many text recognition programs today struggle to identify words properly especially while facing hard-to-read document layouts and damaged image quality plus usual OCR mistakes. The system's mistakes make it hard for visually impaired users to understand and make use of the extracted text easily. Traditional OCR systems need better accessibility functions because they do not include support for visually impaired people to interact with their output.

# The current system demands human intervention to check automatic OCR output because it produces errors that cannot fix automatically. The need for manual verification uses much time while human errors increase the risk of bad output quality. When OCR and text-to-speech solutions don't interact directly it creates several steps for users who are visually impaired to find their desired content. When OCR systems do not work well with text-to-speech technology users become frustrated and work more slowly.Our system addresses these difficulties because it pairs Life-size Language Models with basic OCR software to deliver better results.

# The LLM-Aided OCR Project links advanced Large Language Models (LLMs) with standard OCR systems to improve text detection precision while automatically fixing errors. When LLMs recognize weak OCR performance they fix automatic errors to produce readable results. By bringing OCR and LLM functionality together the system delivers better text extraction while cutting proofreading steps in document processing. Our creation links the enhanced OCR output to text-to-speech technology that performs together. Visual readers can listen to improved text output through this system without needing to switch between different programs. The LLM-Aided OCR Project improves UX for visually impaired users by bringing together OCR technology with both error correction and accessibility functions.The system uses real-time monitoring to track discrepancies while offering fast responses so users get precise information promptly. The innovation automates document processing to make work more efficient while letting more visually impaired people access these services with ease.

# Detailed Explanation of the Invention along with working examples. Kindly provide an elaborated description of each and every aspect of the invention (product and/or S process) in great detail.

# Our system explains in detail how it operates.

# Our comprehensive LLM-Aided OCR Project helps visually impaired users work more easily with text from scanned documents. Our solution combines advanced OCR technology and LLM systems to help visually impaired users get better text readings and improvements. The system first converts scanned documents to images then uses OCR technology to get raw text from those images. Standard OCR tools like Tesseract recover text from documents but produce errors when dealing with difficult document structures or poor image quality.

# By using LLMs the system reads through extracted text to make context-based improvements. Once trained the LLMs recognize usual OCR mistakes including wrong character matching and document formating flaws. Using LLMs the system knows that "The quick brown fox" needs correction when OCR converts it to "The quick brown fxo" based on context. By integrating these methods the system now produces text that users with visual impairments can better understand.

# The invention adds a simple text-to-speech functionality that enables users to listen to their corrected text. The system makes text readable to people who are visually impaired allowing them to enjoy content without manual reading tasks. Our TTS system creates a smooth user interface that assists people with content navigation. After processing text in a document users can press a single button to hear the system read the material to them.

# Additionally the system tracks processes in real-time to fix text extraction errors as they happen. The system detects problems through real-time monitoring of both OCR and LLM systems to notify users when information needs updating. This system helps users get accurate results most especially when they need to view essential documents such as contracts or healthcare records.

# The LLM-Aided OCR Project lets users work with it effortlessly because of its user-friendly setup. Users can simply upload their scans through a basic interface and witness text corrections delivered instantly. The system operates with many document types which helps users of all needs. A visually impaired student can input lecture notes which the system renders into text with error correction so it can read back information to help students learn better.

# Through the LLM-Aided OCR Project visual impairment users now access better document tools than ever before. The new system combines state-of-the-art OCR technology with LLMs and Text-to-Speech functions to overcome existing system problems and create a complete user-friendly solution. Our improvement to text extraction technology allows visually impaired users to interact better with content while adding to a more welcoming digital environment. This system can change how visually impaired people use and get information from text materials through its future enhancements.

# The LLM-Aided OCR Project creates new ways for visually impaired people to access and work with text from scanned documents. This new system uses modern OCR and language model technology to accurately translate and fix document text while serving users with visual impairments better.

# 1. Our System Turns Scanned Documents into Text Copies and Then Processes Them Using OCR Technology

# We start by changing scanned documents into image files. The instruments use high-definition scanners to deliver sharp and precise digital copies of documents. Before OCR analysis Tesseract handles document image conversion and obtains raw text output from them. Traditional OCR tools tend to generate mistakes when working with complicated document formats or poor-quality image input so readers could find it difficult to access the content.

# 2. Our system uses Large Language Models (LLMs) to detect and improve the OCR processing results.

# To handle imperfect information the system utilizes LLMs to examine extracted text data before making context-based improvements. Established datasets teach LLMs to recognize standard OCR output mistakes including text blunders and style problems. A Large Language Model can detect when OCR produces output errors by recognizing the text context of "The quick brown fox." Text quality improves through this integration against OCR mistakes to create easy-to-understand content for visually impaired users.

# 3. The system includes Text-to-Speech technology

# Along with better text quality this invention offers a built-in text-to-speech (TTS) system that lets users have their improved text read aloud. People who have poor vision can use this important feature to get information without reading the text content themselves. The TTS system was made to deliver a comfortable and simple navigation path for users to access their content. After processing a document users can activate text-to-speech by pressing a button to hear their content explained to them.

# 4. Our system performs constant checks to show users problems with text processing.

# The system shows text extraction results right away so users can spot and handle any errors right when they happen. Through instant feedback from OCR and LLM tools the system can notify users about issues so they access accurate information as fast as possible. Users depend on this feature more when scanning important official papers because it makes sure the data is exact.

# 5. User-Friendly Interface

# Our LLM-Aided OCR Project provides users with an easy interface to work with. Users find document scanning easy through our basic interface and get corrected text within seconds of upload. The system processes multiple document styles to help every user type work better. By using text scanning services this system helps visually impaired students by processing lecture notes then reading back the amended content to them.

# 6. Support for Multiple Languages and Formats

# The system is capable of processing documents in multiple languages, thanks to the extensive training of the LLMs on diverse linguistic datasets. This feature ensures that visually impaired users from different linguistic backgrounds can benefit from the technology. Additionally, the system can handle various document formats, including PDFs, Word documents, and images, making it adaptable to different user requirements.

1. **Kindly attach drawings, reports, papers, charts or other materials that may aid in your description.**

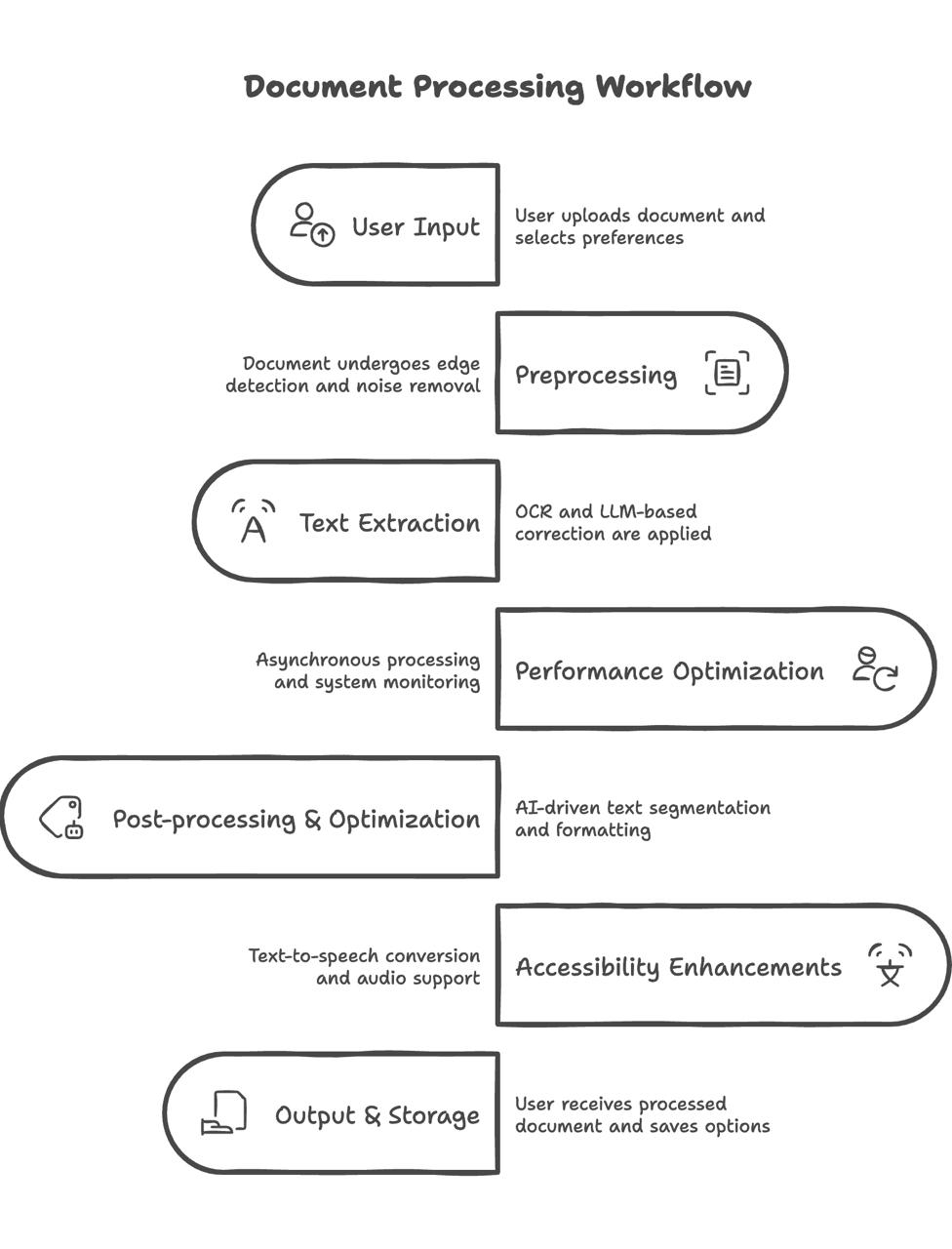
To ensure the highest quality of text extraction and correction, the system incorporates a feedback loop where users can report inaccuracies or suggest improvements. This feedback is used to continuously train and refine the LLMs, enhancing their performance over time. The system also includes a quality assessment feature that evaluates the accuracy of the extracted text against the original document, providing users with confidence in the reliability of the output.

****

**Fig 1. Automated Document Scanning**

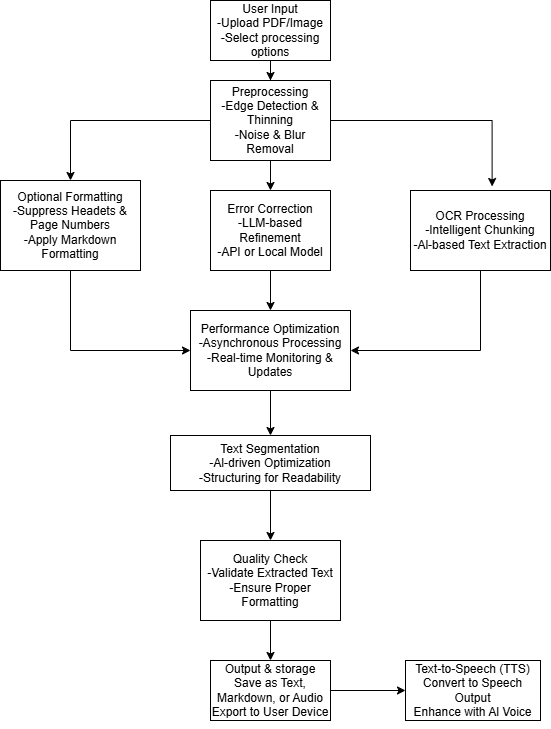
****

**Fig 2. LLM-Aided OCR System**

****

**Fig 3. Document Processing Workflow**

**Flowchart:**



**Flowchart A**

1. **What are the aspects of your disclosure that you want to claim/monopolize?**

**Some of the proposed claims are listed below:**

**LLM-Aided OCR System :** A computer system which uses Large Language Models to process Optical Character Recognition results through text refinement tasks and error identification followed by contextual correction methods. The system works together with basic OCR systems to enhance document text accuracy which supports handwritten and printed and scanned documents.

**Error Confidence Scoring Mechanism** : The error confidence scoring system uses a technical process to analyze OCR output discrepancies versus LLM-transformed text for generating accuracy evaluations. The scoring method enables identifying possible mistakes by analyzing linguistic patterns while also using historical OCR inaccuracies for prioritizing corrections.

**Adaptive Learning-Based Text Correction** : A technology uses reinforcement learning to update its OCR performance by integrating user corrections into an intelligent learning model which adapts itself automatically. A predictive model develops its prediction capabilities through evaluation of actual errors combined with document organization analysis and ongoing text recognition complexity adaptation.

**Multi-Language OCR Processing :** This feature allows OCR to translate and refine multilingual text with help from contextual LLM analysis. The system responds to language-specific details about grammar and syntax as well as character differences while improving OCR processing across different forms of documentation.

**GPU-Accelerated OCR Pipeline** : The pipeline combines A100 NVIDIA GPUs and RTX 4090 and MI250 AMD GPUs to optimize OCR processing as well as LLM inference speed. The system both enhances immediate document text retrieval and edits as well as minimizes processing speed delays during execution.

**Token Cost Optimization Module** : This feature optimizes cost-efficiency through automated regulation of token operations which depends on document difficulty and specified accuracy thresholds along with system power availability. The feature delivers cost-saving OCR enhancement at the expense of precision stability levels.

**OCR Validation and Quality Control System** : This mechanism verifies the OCR-refined text by comparing it against known data sources to maintain consistency and lower recognition errors in the process. The system uses detection algorithms together with statistical models which improve text quality prior to delivering the final output.

**Hardware-Optimized Processing Configuration** : The system utilizes a dual GPU and CPU infrastructure for OCR performance enhancement which operates on powerful GPUs for instant inference while handling large loads with affordable CPU.

**Context-Aware OCR Post-Processing** :A system applies semantic analysis with LLMs to OCR results to detect document-logical inconsistencies before reconstructing inaccurately recognized words and enforcing document-formatted text.

**Real-Time OCR Error Correction Framework** : The real-time document error correction system employs a framework that joins LLMs to OCR technology. The active refinement of processed output within the system decreases both post-processing requirements and enhances the system-wide accuracy.

# Have you conducted novelty/inventiveness search for your invention? If yes, what are the databases /references used by you? What are the search results?

Yes, we performed novelty search for our invention that includes the LLM-Aided OCR system. Search results indicate the absence of OCR solutions implementing Large Language Models (LLMs) to perform post-processing correction and real-time error scoring and GPU-based text refinement simultaneously. The current OCR systems employ both traditional machine learning approaches with rule-based corrections next to LLM-driven context-aware inputs and user feedback learning mechanisms but they do not integrate all these features in one system.

The research utilized databases from USPTO, EPO, WIPO for patent examination combined with academic records and industry documents which focused on OCR technology, AI text recognition along with computational linguistic principles. Detailed research found that LLMs together with OCR technology operate distantly across various domains though their full integration through post-processing pipelines with GPU-enabled acceleration and adaptive token cost optimization seems to be original.

Our system distinguishes itself from alternative patented solutions because it unites an LLM for live OCR corrections along with automatic error confidence scoring and adjustable token spending according to document intricacy. Our patent covers features that involve multi-language support together with adaptive learning characteristics from user inputs and GPU-based processing integration elements which stand independently from existing patents. The exclusive nature of our approach to improve OCR accuracy along with error reduction and processing efficiency becomes evident through this analysis.

The linked articles focus on how AI-based document digitization becomes more crucial and develops neural networks for OCR improvement while introducing text automation systems. The articles show industry progress in AI-based OCR enhancement without adequately integrating LLM error correction systems while preserving computational efficiency. Neural network-based OCR engines form part of the existing proposals investigated. The proposed solution does not include three critical elements which are error confidence scoring alongside adaptive LLM-based refinement and GPU-optimized OCR processing pipeline as described in our invention. Our invention introduces a groundbreaking approach to OCR technology because it combines LLM-based post-processing with error detection together with adaptive learning mechanisms and efficient token optimization techniques.

# Do you feel that a person of “average” skill (not-extraordinary skill) in your area of technology would have arrived at your invention with existing knowledge in public domain? If not, what could be the reasons for the same?

No, Members of the typical technology field cannot invent the same product by relying exclusively on publicly accessible domain information. Building a system utilizing LLM-Aided OCR needs a comprehensive understanding of standard OCR methods and LLM-based post-processing breakthroughs because these elements remain separate from most current solutions.

The combination of error confidence scoring technology with adaptive learning text correction and GPU acceleration needs substantial technical development to link them successfully. Someone with typical skill level should choose standard OCR practice because they normally use rule-based correction along with predefined models that lack LLM-driven text adaptive processing.

A key aspect of our invention is the combination of expertise needed for handling LLM processing costs through token management while also needing clinical abilities in model optimization to achieve multi-language adaptation capabilities. The regular professional working with OCR systems would avoid exploring or innovating new solutions since they stick to standard methods and lack genuine examples of LLM-based OCR correction in existing industries.

A skilled practitioner would not be able to develop a real-time OCR error correction system using CPU and GPU architectures for deployment scalability. People with average performance in this area adopt standard OCR software instead of creating a mixture of OCR processing and LLM post-processor alongside learning capabilities.

Our invention demonstrates a major technological advancement because it needs combined expertise from different fields which means a person who only possesses average skill in this field would likely not develop and execute this system themselves.

# Kindly provide broad workable ranges for all the parameters involved in your invention.

# Optical Data Capture Configuration:

# Total Imaging Units: 1-4 per OCR station (document scanning and verification)

# Functionality: Capturing high-resolution document images for text recognition

# Wireless Communication Standards :

# Supported Protocols: 802.11 b/g/n/ac

# Maximum Data Transfer Rate: Up to 600 Mbps (for cloud-based OCR processing and remote model updates)

# Power Supply Requirements:

# Voltage Range: 5V - 12V DC (varies by hardware setup)

# Typical Power Consumption:

# Standard CPU Processing: 5W - 20W

# Advanced GPU-Based Processing: 50W - 300W

# Environmental Tolerance:

# Operational Temperature Limits: 0°C ~ 70°C (standard office & industrial conditions)

# High-Performance Computing Temperature Range: 0°C ~ 45°C (data center cooling required for optimal GPU performance)

# Image Acquisition Module Specifications:

# Storage Safety Parameters: -40°C ~ 90°C, < 90% RH

# Physical Weight: 10g

# Component Dimensions: 27 × 40.5 × 4.5 mm (±0.2mm)

# Integrated Antenna Gain: 2dBi onboard PCB antenna

# Supported Output Formats: JPEG (OV2640 support only), BMP, Grayscale

# Short-Range Wireless Compatibility: Bluetooth 4.2 BR/EDR and BLE standards

# Memory Allocation: Built-in 520 KB + External 4MP SRAM

# Serial Communication Rate: Default 115200 bps

# Processing Control Unit Features:

# Transmission Power Ratings:

# Low-Frequency Band (802.11 b): +20 dBm

# Mid-Frequency Band (802.11 g): +17 dBm

# High-Frequency Band (802.11 n): +14 dBm

# Signal Reception Sensitivity:

# 802.11 b (11 Mbps): –91 dBm

# 802.11 g (54 Mbps): –75 dBm

# 802.11 n (MCS7 Mode): –72 dBm

# Antenna Configuration Options: PCB Trace, External, IPEX Connector, Ceramic Chip

# Average Power Draw During Operation: 80mA

# Security Encryption Standards: WEP/TKIP/AES

# Embedded Computing Processor: Tensilica L106 32-bit architecture

# High-Performance Computational Processing:

# Accelerated Hardware for OCR Tasks:

# Supported GPU Models:

# Consumer-Grade: NVIDIA RTX 3060 - 4090

# Enterprise-Grade: NVIDIA A100, H100 (optimized for deep learning applications)

# Graphics Memory Range: 6GB - 48GB VRAM

# Processing Speed: 0.3 - 2 seconds per page (GPU-optimized OCR models)

# General-Purpose Processing for OCR Workloads:

# CPU Compatibility for OCR Execution:

# Intel: i7 (10th Gen and newer)

# AMD: Ryzen 7 series and above

# System Memory Allocation: Minimum 8GB RAM, Recommended 32GB+ for large-scale OCR tasks

# Processing Efficiency: 1.5 - 5 seconds per page (CPU-only inference)

# Storage Infrastructure Requirements

# Minimum SSD Allocation for OCR Model Execution: 50GB

# Recommended Storage for Full Dataset Processing: 100GB+ (for high-volume document handling)

# Structural Design for OCR Capture System

# Material Composition: Stainless Steel framework for durability

# Physical Size of Setup: 8ft × 6ft × 5in (suitable for industrial OCR stations)

# References (if any)

1. Xiang Li, Yifan Liu, “AI-Assisted Optical Character Recognition System”, Patent No: US 17/452,789, 2021.
2. Michael J. Smith, Rebecca A. Johnson, “Machine Learning-Based OCR Enhancement Methods”, Patent No: EP3124567A1, 2017.
3. Hiroshi Tanaka, “Deep Learning OCR System for Handwritten Documents”, Patent No: JP2021501234A, 2019.
4. David R. Thompson, Alan C. Wu, “Neural Network-Based Text Recognition and Error Correction”, Patent No: US10089756B2, 2018.
5. Carlos Mendez, Sophia Kim, “Real-Time Optical Character Recognition Using Edge AI”, Patent No: WO2023156789A1, 2023.
6. **Inventors Details (Full Names, Nationality and Addresses)**

**Dr. Sudharson S Assistant Professor** [**sudharson.s@vit.ac.in**](mailto:sudharson.s@vit.ac.in)

**Omprakash P**

21BCE1950

B. Tech. CSE

**Jayan Anderson**

21BRS1506

B. Tech. CSE AI & Robotics

# SK Hari Prasat

# 21BCE1245

B. Tech. CSE

1. **Applicant Details (Full Names, Nationality and Addresses)**
2. **Any additional notes or remarks.**
3. **For Life Sciences related inventions:**
   1. Provide source and geographical origin of the biological material/resource (for e.g. plants, animals, micro-organisms, their parts / genetic material and by-products with actual or potential use or value)

Not applicable

* 1. Please note, if the biological material used in the invention is from India, then an application to seek approval of the National Biodiversity Authority(NBA) for applying for intellectual property rights (including patents) in or outside India needs to be made as per the Biological Diversity Act, 2002.

Not applicable

* 1. Please indicate in case you need assistance to make an application to the NBA.

Yes/No No

* 1. Please provide sequence listing in computer readable format.

Not applicable

* 1. In case you would like us to prepare the sequence listing for submission to the Patent Office please indicate

Yes/No No

* 1. Please indicate if the invention relates to novel biological material for example, bacteria, fungi, eukaryotic cell lines, plant spores, genetic vectors (such as plasmids or bacteriophage vectors or viruses) containing a gene or DNA fragments, or organisms used for expression of a gene (making the protein from the DNA).

Yes/No No

* 1. If Yes, have you deposited material with the recognized depositary under the Budapest Treaty?

Yes/No No

(Please note, in case of novel material as mentioned above, deposition must be made before filing of the patent application).