www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 694-699

e-ISSN : 2583-1062
Impact Factor : 7.001

FORNOVA – DOCUMENTS WITH AI

Shrihari Suresh Kasar[1], Rohit Pramod Khadangle[2], Utkarsha Bajirao Kakulte[3], Ishwari Sandip Gamne[4], Rohini D. Bhaskar[5]

[1,2,3,4]Student, Computer Engineering, Matoshri College Of Engineering And Research Centre, Nashik, Maharashtra, India.

[5]Assistant Professor, Computer Engineering, Matoshri College Of Engineering And Research Centre, Nashik, Maharashtra, India.

ABSTRACT

An inventive use of artificial intelligence to improve, automate, and streamline document management procedures across industries is represented by FORNOVA-Documents with AI. FORNOVA intelligently classifies, extracts, summarizes, and validates data from a variety of document types, including invoices, contracts, reports, and more, by utilizing machine learning, natural language processing (NLP), and computer vision. The system significantly lowers manual labor, minimizes human error, and speeds up decision-making by incorporating AI technologies. It facilitates smooth interoperability across platforms and departments by supporting multilingual and multiformat document handling. Through constant learning from user interactions, the system guarantees flexibility and gradually increases accuracy. FORNOVA offers a scalable, secure, and intelligent solution that turns conventional document workflows into dynamic, AI-driven processes as businesses deal with increasing amounts of unstructured data. This innovation positions itself as a pillar for the digital transformation of document management by improving compliance, data accuracy, and strategic insights in addition to operational efficiency. Additionally, by providing modular AI components that can be tailored to meet organizational needs, FORNOVA advances the expanding field of Intelligent Document Processing (IDP). The system ensures accessibility and real-time collaboration by integrating with cloud storage and APIs. All things considered, FORNOVA-Documents with AI offers a comprehensive digital framework that substitutes a self-learning, automated, and effective ecosystem for conventional, manual document handling, providing a contemporary method of information management that is intelligent, transparent, and accurate.

Keywords: Natural Language Processing (NLP), Machine Learning (ML), Intelligent Document Processing (IDP), Unstructured Data, Document Automation, And Artificial Intelligence (AI).

1. INTRODUCTION

Information is now one of the most important resources for businesses in the contemporary digital economy. Every business, whether in the hospitality, financial, medical, educational, or logistical sectors, depends on an ongoing flow of documents to run efficiently. These documents come in a variety of formats, including contracts, invoices, reports, forms, and compliance records. But handling this enormous and expanding amount of unstructured data is a constant struggle. Sorting, data entry, and verification are all examples of manual document handling that is laborious, error-prone, and frequently results in inefficiencies and compliance problems. Because of this, businesses are looking for intelligent systems that can automate document-related tasks while preserving security, scalability, and accuracy. Every day, Fornova, a multinational technology company that works in the analytics space for travel and hospitality, handles a lot of complicated documents. These could include regulatory documents, client communications, competitor benchmarking reports, and audits of hotel rates. These procedures historically depended on human analysts to interpret and classify data, which led to high operating costs, delays, and inconsistencies. Furthermore, it can be challenging to turn raw data into intelligence that can be put to use because important insights are frequently concealed within unstructured text. In light of these difficulties, the project "FORNOVA – Documents with AI" intends to transform document processing through the introduction of an intelligent automation framework driven by NLP, ML, and AI. A key component of Industry 4.0, where automation, data sharing, and intelligent analytics rule corporate operations, is the digital transformation of document management systems. Conventional document management systems (DMS) have limited capacity for intelligent data interpretation and are mainly concerned with storage and retrieval tasks. As a result, classification, extraction, and summarization tasks still require human input from organizations. Such reliance adds subjectivity and error risk in addition to lengthening processing times. It is clear that a more dynamic, self-learning, and context-aware system is required. The basis for accomplishing this change is artificial intelligence. Without explicit programming, systems can automatically learn from previous examples and adjust to new document types thanks to machine learning algorithms. The system can interpret human language, derive semantic meaning, and condense important information thanks to natural language processing. Optical Character Recognition (OCR) is one computer vision technique that turns handwritten or scanned text into

machine-readable information. Intelligent Document Processing (IDP), a new paradigm that manages documents end-to-end by fusing automation and cognitive understanding, is the result of these technologies' integration. By providing a single platform for document classification, data extraction, summarization, validation, and intelligent search, the FORNOVA system exemplifies these ideas. Its modular architecture maintains high adaptability for a variety of document formats and languages while enabling scalability across multiple industries. The solution uses machine learning models to recognize document types and extract pertinent fields, OCR to digitize handwritten and printed inputs, and NLP to comprehend content. In addition, the system uses cross-referencing and rule-based validation methods to guarantee data integrity, which is essential for audit and compliance procedures. Data fragmentation is a major problem in document management. Because documents are frequently dispersed throughout various departments and storage systems, retrieval is inconsistent and inefficient. This is addressed by FORNOVA, which enables smooth system interoperability through centralized cloud storage and metadata tagging. Natural language queries can be used by users to search documents, removing the need for strict keyword-based searches and facilitating intuitive interaction. By significantly cutting down on the amount of time needed to find specific information, this intelligent retrieval system enhances the effectiveness of decision-making. The use of AI-driven document processing has a number of quantifiable benefits from an operational standpoint. First, by automating repetitive tasks like data entry and report generation, it lowers operational costs and manual workload. Second, it improves accuracy and consistency because, with the right training, AI models can function with little human bias. Third, by keeping thorough logs and audit trails for each processed document, it facilitates adherence to legal and regulatory requirements. Last but not least, the solution's scalability enables businesses to handle thousands of documents every day without experiencing performance issues. IDP solution research and industry adoption have accelerated in recent years. AI-based document processing platforms were first introduced by companies like UiPath, ABBYY, and Kofax; nevertheless, these solutions frequently have expensive licensing fees and little room for customization. The FORNOVA-Documents with AI initiative set itself apart by emphasizing an open, adaptable, and affordable architecture that can be used by both large corporations and small-to-medium businesses (SMEs). The system guarantees affordability and scalability by integrating cloud integration (AWS, Azure, or GCP) with open-source AI libraries like TensorFlow and PyTorch. Additionally, the FORNOVA framework includes a feedback-driven learning loop that retrains the AI models based on user corrections and approvals, leading to ongoing improvement over time. This method creates a hybrid intelligence system that adapts to the demands of the organization by bridging the gap between automation and human expertise. By using neural machine translation to support documents in English, Hindi, and other major languages, the platform's multilingual capability further broadens its applicability across geographies and industries. A key component of the FORNOVA system's design is security and privacy. The system uses encryption for data transmission and storage because documents may contain private or sensitive financial information. Only authorized individuals are able to view, modify, or approve processed data thanks to role-based access control, or RBAC. Transparency and traceability are essential in regulated environments, and audit trails and tamper-proof logs offer both.

The FORNOVA project's introduction essentially highlights the increasing demand for intelligent automation in document management. The system turns static, paper-based processes into dynamic, self-improving workflows by utilizing AI, NLP, and ML. In addition to advancing technology, the research tackles practical issues like operational inefficiency, data inaccessibility, and compliance complexity.

2. LITERATURE SURVEY

- AI-Powered Document Automation by Kumar et al. (2021)

Categorized and extracted data from digital documents “using machine learning and natural language processing techniques”. The study highlighted the effectiveness of intelligent systems in handling massive amounts of data by showing that AI-based automation dramatically lowers manual labor and increases document handling accuracy.

- According to Mehta (2020), Intelligent Document Processing (IDP)

Automated text comprehension and information retrieval “by combining optical character recognition (OCR) and natural language processing (NLP)”. The study demonstrated that structured extraction from unstructured documents is made possible by IDP systems, which served as the conceptual foundation for FORNOVA's seamless document analysis integration of OCR and NLP.

- Deep NLP Methods for Business Document Analytics by Sharma & Patel (2022)

Used transformer models, like BERT, to classify and summarize documents. Their methodology guided FORNOVA's design for deep NLP-based auto summarization and keyword detection, improving contextual understanding and summarization accuracy for intricate business reports.

- Gupta (2021): Summarizing Documents with Neural Networks

Used sequence-to-sequence neural architectures to condense lengthy textual data into manageable summaries. Through accurate summarization, the study improved the efficiency of document retrieval and had an impact on FORNOVA's automatic summary generation module for legal and business reports.

3. METHODOLOGY

The goal of the proposed FORNOVA-Documents with AI project is to create an intelligent, end-to-end document management system that automates document handling for businesses by utilizing machine learning, artificial intelligence, and natural language processing. Classification, data extraction, summarization, translation, and validation are all done within a single framework by the system, which is made to process unstructured and semi-structured data effectively. System analysis, architecture planning, AI model design, and secure data flow implementation are all part of the project's structured methodology.

3.1 SYSTEM OVERVIEW:

The fundamental idea behind FORNOVA is to use an intelligent, adaptable AI platform that can learn from experience to replace manual document processing. AI algorithms will be used to automatically analyze documents that users upload in a variety of formats, including scanned copies, text files, PDFs, and images.

3.2 THE SYSTEM SEEKS TO:

1. Determine the document type (report, contract, invoice, etc.).
2. Key data fields, such as names, dates, amounts, and clauses, should be extracted.
3. Create summaries of lengthy documents automatically.
4. Use pre-established rules to validate the extracted data.
5. If necessary, translate text into several languages.

FORNOVA will provide a comprehensive solution that lowers manual intervention, boosts productivity, and enhances data reliability by combining these tasks into a single pipeline.

3.3. SYSTEM ARCHITECTURE:

Layer	What It Does	Technologies / Tools Used
User Interface Layer	This is what users and administrators interact with. It allows login, viewing dashboards, uploading documents, and browsing content easily.	React.js, HTML, CSS, JavaScript
AI Processing Layer	This layer handles the “smart” part: it extracts text from documents, classifies content, summarizes information, and even translates text when needed.	Python, TensorFlow, PyTorch, spaCy, NLTK
Validation and Storage Layer	Ensures that the extracted information is accurate and correctly formatted before storing it safely in a database.	MySQL, MongoDB, Flask APIs
Cloud and Security Layer	Keeps your data safe by encrypting transfers, managing cloud storage, and controlling access based on user roles.	AWS, Google Cloud, Azure, JWT, Firebase Auth

4. MODELING AND ANALYSIS

The following describes the suggested FORNOVA-Documents with AI workflow:

1. User authentication: A secure authentication portal is used by users or administrators to register or log in. Only authorized personnel are able to carry out administrative tasks thanks to role-based access.

2. Uploading Documents: Users can upload documents in Word, PDF, and image formats. Metadata such as the uploader ID, type, and date of upload are attached to every document.

3. Pre-processing Step: To eliminate noise, correct orientation, and get ready for AI processing, uploaded files are pre-processed using OCR and image enhancement.

4. Analysis Based on AI: Several modules are used to analyze the document:

5. Classification: Machine learning models determine the type of document (e.g., financial, legal, healthcare, etc.).

6. Names, numbers, and addresses are among the entities that are extracted by NLP models.

7. Review and Approval: The administrator reviews the AI-generated output, approves or requests corrections, and flags any anomalies for reprocessing.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 694-699

e-ISSN : 2583-1062
Impact Factor : 7.001

8. Storage and Retrieval: Verified documents and extracted information are stored securely in the database, which supports intelligent search and filtering.

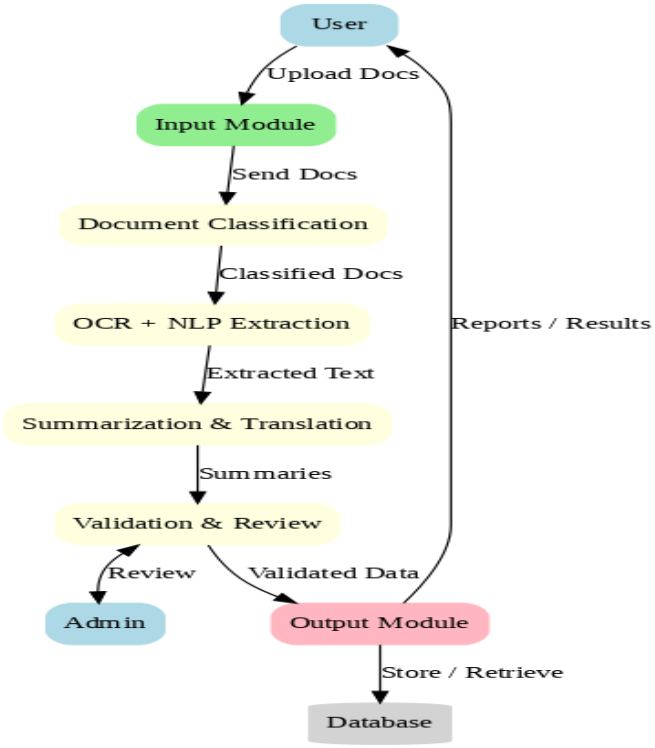

Fig 4.1: Workflow Diagram of FORNOVA

4.1 EXPECTED OUTCOMES

The FORNOVA system is anticipated to:

1. 80–90% of manual document handling tasks should be automated.
2. For common document types, attain high extraction accuracy (≥90%).
3. Compared to manual workflows, cut the processing time per document by more than 70%.
4. Offer user-level authentication for safe, cloud-based document access.
5. Allow searchable archives and real-time summaries to facilitate quick data retrieval.

The ultimate goal is to develop an intelligent, scalable, and adaptable framework that can grow across several industries and change with user data.

5. RESULTS AND DISCUSSION

5.1 PROTOTYPE DESIGN

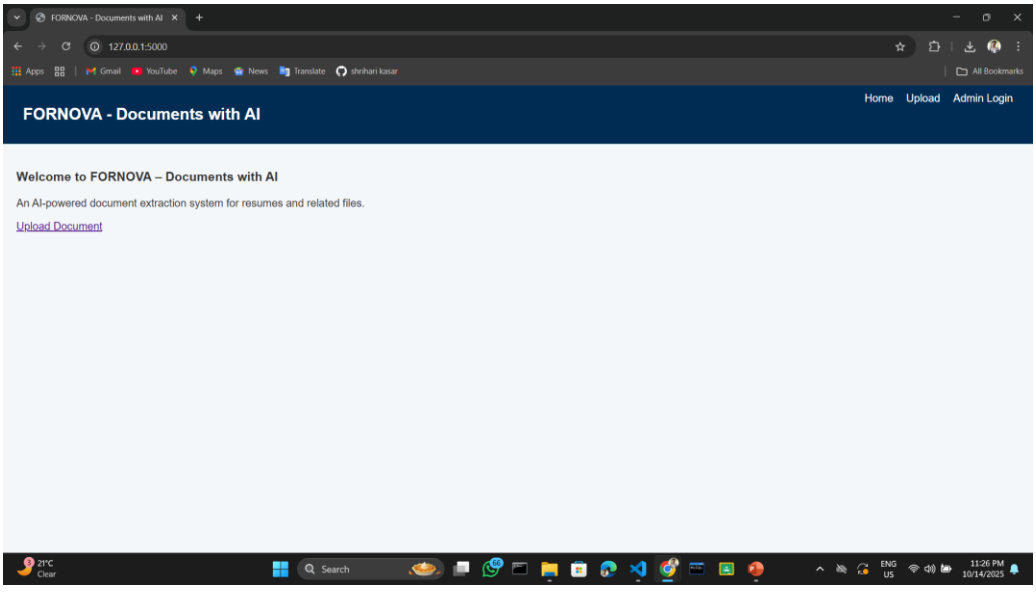

Fig 5.1: User Interface

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 694-699

www.ijprems.com
editor@ijprems.com

e-ISSN : 2583-1062
Impact Factor : 7.001

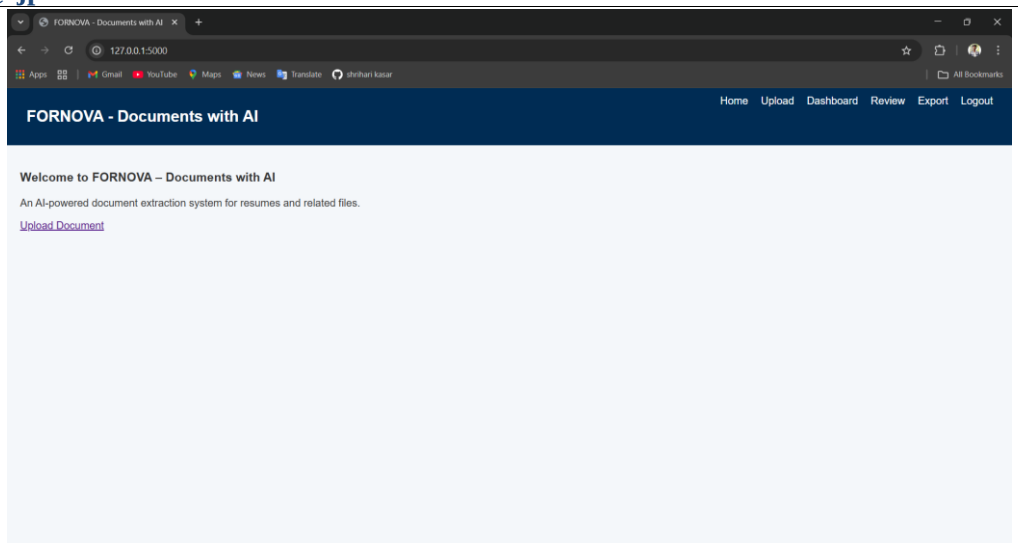

Fig 5.2: Admin Dashboard

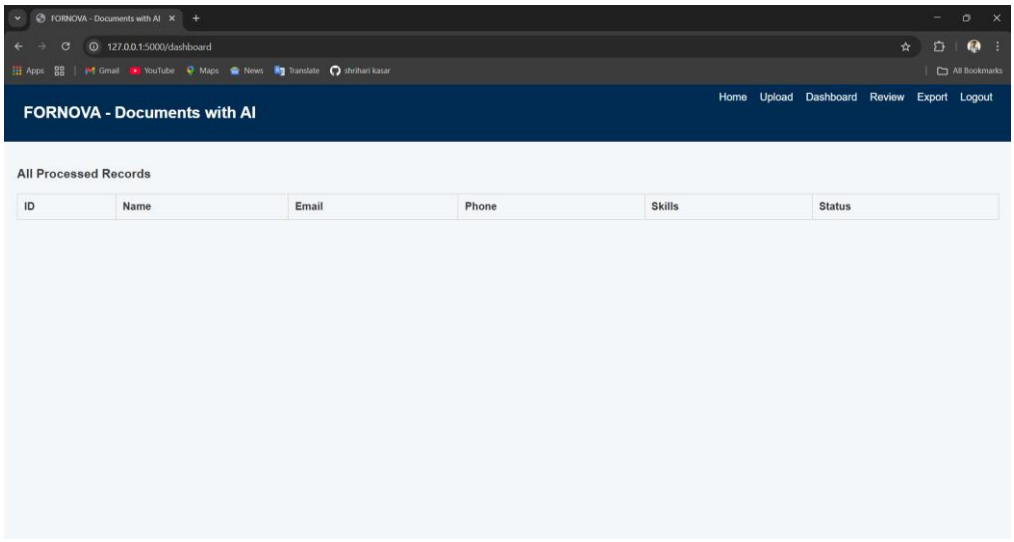

Fig 5.3: Extracted Data (Review Dashboard)

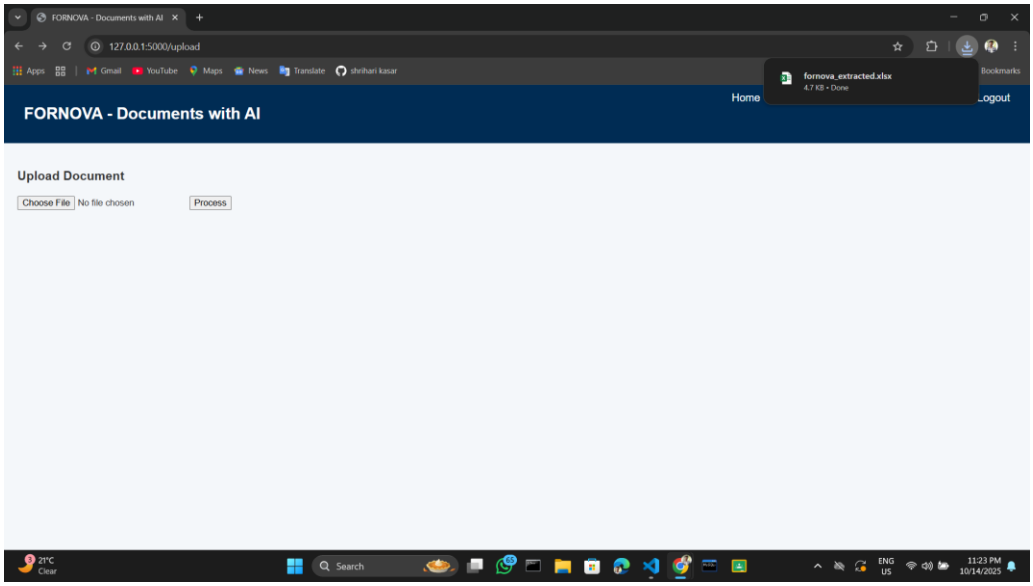

Fig 5.4: Upload And Save To Excel By Export Button

5.2 FUTURE ENHANCEMENTS

Following testing and successful deployment, the system can be extended to include:

1. Integration with speech-to-text APIs for voice-based document queries.

2. Sophisticated AI feedback loops for ongoing development and self-learning.
3. Blockchain-based data validation for tamper-proof document storage.
4. Multilingual support for global applicability across different regions.

5.3 CHALLENGES AND LIMITATIONS:

1. Managing Unstructured Documents: Handwritten or poorly formatted documents may result in less accurate text extraction.
2. Domain-Specific Language: Regional linguistic subtleties or technical jargon may be misinterpreted by AI models.
3. Scalability Issues: Processing a lot of documents can cause performance issues and resource consumption to rise.
4. Data Security and Privacy: Strict adherence to regulations is necessary for the cloud storage of sensitive data, even when encryption is used.
5. Model Maintenance: To maintain accuracy, AI models must be retrained on a regular basis, which takes time and money.

Integration Challenges: It can be difficult to guarantee smooth communication between the cloud, storage, AI, and user interface layers.

6. CONCLUSION

A smooth, safe, and intelligent document management experience is guaranteed by the multi-layered architecture. While the AI Processing Layer greatly reduces manual labor by automating text extraction, classification, summarization, and translation, the User Interface Layer offers an intuitive platform for administrators and users. The Cloud and Security Layer use encryption and role-based access controls to protect data, while the Validation and Storage Layer ensure data accuracy and dependability. When combined, these layers produce a strong system that is effective, safe, and expandable, satisfying the needs of contemporary document management and AI-powered processes.

7. REFERENCES

[1] Bengio, Y., Goodfellow, I., and Courville, A. (2016). MIT Press, "Deep Learning."
[2] F. Chollet (2018). Manning Publications, "Deep Learning with Python."
[3] Bird, S., Loper, E., & Klein, E. (2009). Python-based natural language processing. Media O'Reilly.
[4] TensorFlow (2025). TensorFlow is an open-source, end-to-end machine learning platform. taken from https://www.tensorflow.org.
[5] PyTorch, 2025. Python Tensors and Dynamic Neural Networks with PyTorch. taken from the website https://pytorch.org
[6] AWS Documentation, 2025. Best Practices for Amazon Web Services Security. taken from https://aws.amazon.com/security
[7] MongoDB (2025). Documentation for MongoDB. taken from https://www.mongodb.com/docs
[8] MySQL (2025). Reference Manual for MySQL 8.0. taken from https://dev.mysql.com/doc
[9] Firebase (2025). Guide to Firebase Authentication. taken from https://firebase.google.com/docs/auth
[10] spaCy (2025). Python-based natural language processing with industrial strength. taken from the website https://spacy.io
[11] NLTK (2025). Documentation for the Natural Language Toolkit. taken from https://www.nltk.org