ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

SoftLocker: Document Management Through Automated Classification and Cloud Integration

Deepa A. V.[1]; Premalatha[2]

[1]Department of Computer Science and Engineering RVS Technical Campus-COIMBATORE Kumaram Kottam Campus, Kannampalayam, Coimbatore-641402
[2]Department of Computer Science and Engineering RVS Technical Campus-COIMBATORE Kumaram Kottam Campus, Kannampalayam, Coimbatore-641402

Publication Date: 2026/05/20

Abstract: SoftLocker is a cloud-based document management web application designed to automate the organization and storage of documents using advanced Artificial Intelligence (AI) technologies. Traditional document management methods often involve manual sorting and storage, which can lead to ineffi-ciencies, misplacement of files, and security risks. SoftLocker addresses these challenges by providing an intelligent system that simplifies document handling and improves the overall efficiency of managing digital documents.

The system incorporates technologies such as Optical Charac-ter Recognition (OCR), keyword detection, and AI-driven doc-ument categorization to process uploaded documents automati-cally. Users can upload scanned documents or images through the application, after which the system extracts text using OCR and analyzes the content to identify important keywords. Based on this analysis, machine learning algorithms categorize the documents into appropriate groups, ensuring organized storage with minimal human intervention. SoftLocker securely stores documents in the cloud, enabling users to access their files anytime and from anywhere while maintaining strong security measures. The application also integrates machine learning models that improve document classification accuracy over time as more data is processed. By offering secure storage, easy accessibility, and intelligent document organization, SoftLocker provides individuals and businesses with a reliable and user-friendly solution for managing digital documents and reducing the risk of data loss, especially in critical situations such as natural disasters.

Keywords: *Natural Language Processing, Optical Character Recognition.*

How to Cite: Deepa A. V.; Premalatha (2026) SoftLocker: Document Management Through Automated Classification and Cloud Integration. *International Journal of Innovative Science and Research Technology*, 11(5), 692-697. https://doi.org/10.38124/ijisrt/26may232

I. INTRODUCTION

Organizations and people are continuously creating and managing massive amounts of documents in both digital and physical media in the current digital era. As the amount of information continues to increase, managing these papers effectively has grown more difficult. Conventional document management techniques frequently depend on manual classifi-cation, physical storage, or rudimentary digital filing systems, which can result in problems like lost files, ineffective search-ing, limited scalability, and possible security flaws. These drawbacks emphasize the necessity for a clever, automated system that can streamline document management while guar-anteeing safe storage and convenient access.

In order to overcome these obstacles, SoftLocker presents a cutting-edge document management system that makes use of cutting-edge machine learning and artificial intelligence (AI) technologies to optimize the complete document lifecycle. Document scanning, text extraction, classification, and storage are only a few of the document processing steps that the sys-tem is intended to automate. SoftLocker uses optical character recognition (OCR) to analyze scanned documents and photos, transform the text into a machine-readable format, and recog-nize key words and phrases. The system's AI models then use these extracted keywords to automatically classify documents and arrange them into the proper folders, greatly reducing the need for manual input and lowering the possibility of human error.

SoftLocker's ability to dynamically organize and label saved documents based on the information it detects is another important feature. Users may find files quickly and effectively because to the system's ability to create relevant folder names and classify documents based on their context. This automated system decreases the amount of time spent looking for crucial information and increases productivity. As

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

more documents are processed, the embedded machine learning algorithms gradually increase their categorization accuracy, making the system more dependable and flexible with regard to various file kinds.

SoftLocker also utilizes cloud-based storage infrastructure through Firebase integration, enabling secure and scalable document storage. By storing documents in the cloud, users can access their files anytime and from any device with internet connectivity. This ensures flexibility and convenience, espe-cially for businesses or individuals who require remote access to important documents. Additionally, the use of cloud services improves data reliability by providing backup mechanisms that protect against accidental data loss, device failure, or other unforeseen circumstances.

Security is a fundamental component of the SoftLocker platform. The system incorporates robust encryption techniques to protect stored documents and prevent unauthorized access. Furthermore, role-based access control mechanisms ensure that only authorized users can view or modify sensitive documents. These security measures make SoftLocker suitable for handling confidential information while maintaining user privacy and data integrity.By combining advanced AI-driven automation, secure cloud storage, and a user-friendly interface, SoftLocker provides a powerful and efficient document man-agement solution. The system not only simplifies document organization but also enhances security, accessibility, and pro-ductivity. Through its intelligent design and modern technolog-ical framework, SoftLocker redefines the way documents are handled, making the process faster, safer, and more efficient while allowing users to focus more on their core activities rather than administrative tasks.

II. RELATED WORKS

➢ *Advanced Automated Document Processing Using Optical Character Recognition (OCR)*

This paper focuses on advancing document processing using Optical Character Recognition (OCR) to address the rising volume of handwritten and scanned documents across sec-tors like banking and education. The study aims to reduce manual data entry and improve accuracy through a model that leverages OCR technology to efficiently extract text from scanned documents. Key preprocessing techniques, including skew correction, deblurring, and contrast enhancement, are utilized to enhance character recognition accuracy [1]. The methodology used in ”Advanced Automated Document Pro-cessing Using Optical Character Recognition (OCR)” focuses on enhancing OCR accuracy, particularly for handwritten text. The approach involves pre-processing techniques such as noise reduction, binarization, and contrast enhancement to improve input image quality. Feature extraction techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are employed to recognize complex handwritten patterns. The system integrates machine learning models trained on diverse datasets to enhance recognition accuracy across different handwriting styles. Finally, post-processing techniques, such as language models and spell-checking algorithms, are applied to refine and validate the extracted text. This multi-step process significantly improves the reliability of automated document processing.

- *Advantages:*

Implementing advanced OCR technologies and pre-processing techniques can significantly improve the accuracy of recognizing handwritten text. This enhancement leads to more reliable data extraction from documents, reducing errors associated with manual data entry. Consequently, organizations can experience increased efficiency and productivity in document management processes.

- *Disadvantages:*

Despite the advancements, challenges re-main in achieving high accuracy rates, especially with diverse handwriting styles and document qualities. The implementation of sophisticated OCR systems may require substantial computational resources and initial setup costs. Additionally, continuous maintenance and updates are necessary to adapt to varying document formats and to further improve recognition Capabilities.

➢ *Paperless Paradigm: Intelligent Automation in Document and Record Management*

The paper explores the transformative impact of intelligent automation on document and record management systems. It delves into how automation technologies can streamline processes, enhance accuracy, and reduce the reliance on physical doc- uments. The study highlights the benefits of a paperless environment, including improved efficiency, cost savings, and environmental sus-tainability. By leveraging ad- vanced technologies such as artificial intelligence and machine learning, organizations can achieve more effective and secure management of their records and documents, paving the way for a more efficient and eco-friendly future.[2]

- *Methodology:*

The methodology proposed in *Paperless Paradigm: Intelligent Automation in Document and Record Management* focuses on integrating **artificial intelligence (AI) and robotic process automation (RPA) ** to streamline document and record management. The process begins with **AI-driven algorithms** that handle data extraction, text recognition, and document classification. These algorithms leverage machine learning (ML) and natural language processing (NLP) to identify and categorize docu-ments with minimal human intervention. Meanwhile, **RPA is employed to automate repetitive administrative tasks**, such as data entry, document sorting, and routing. The system also includes cloud-based storage and blockchain for enhanced security and accessibility. By combining AI’s cognitive capa-bilities with RPA’s efficiency, the methodology aims to create a fully automated, paperless document management system that improves workflow efficiency and reduces manual workload. Advantages: Implementing intelligent automation in doc-ument and record man- agement offers several significant benefits. **Efficiency is greatly enhanced**, as AI and RPA accelerates document processing, reducing delays caused by

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

manual handling. **Accuracy improves**, as AI minimizes human errors in data extraction, ensuring better document classification and retrieval. **Cost savings** are another major advantage, as organizations can reduce expenses related to physical storage, labor, and printing. Additionally, the system is **highly scalable**, allowing businesses to manage an increasing volume of documents without requiring additional human resources. Moreover, **automated document tracking and encryption** improve compliance with regulatory standards and enhance data security, ensuring that sensitive records remain protected from unauthorized access.

- *Disadvantages:*

Despite its advantages, intelligent automa-tion in document man- agement presents certain challenges. **High initial investment costs** can be a barrier for many organizations, as AI and RPA implemen-tation require financial resources for software, hardware, and staff training. Additionally, the **complexity of integrating au-tomation** into existing workflows may require companies to restructure their document management strategies, which can be time-consuming and disruptive. **Data security concerns** also arise, as automated systems handling sensitive records need strong encryption and cybersecurity measures to prevent breaches. Furthermore, **over-reliance on technology** can be a risk, as system failures, algorithmic errors, or cyberattacks could disrupt automated workflows, leading to operational downtime and potential data loss.

➢ *Transforming Scene Text Detection and Recognition: A Multi- Scale End-to-End Approach With Transformer Frame-Works*

This paper explores a multi-scale, end-to-end approach with a Transformer framework for detecting and recognizing text in images. It addresses challenges in scene text detection and recognition, including noise, diverse text fonts, and low image quality.The model is evaluated on several datasets (e.g., COCO-Text, SynthText, ICDAR 2017), using precision, recall, and F1 scores. Cross-validation techniques were applied to enhance robustness, and results were analyzed using various metrics to confirm the efficacy of multi-scale and end-to-end approaches.[4] Methodology: The paper **"Transforming Scene Text Detection and Recognition: A Multi-Scale End-to-End Approach with Transformer Framework"** introduces an advanced methodology that integrates text detection and recognition into a single **end-to-end framework**. It lever-ages **multi-scale feature extraction** to improve accuracy in identifying text across various sizes and orientations in natural scenes. The **transformer framework** plays a crucial role by utilizing self-attention mechanisms to efficiently process text in complex backgrounds, improving the model's ability to generalize across different datasets. By incorporating **po-sitional encoding and multi- layer stacking**, the approach ensures a more robust understanding of text structure, allowing for better handling of irregularly shaped and distorted text. This unified system reduces computational overhead compared to traditional two-step methods, making it more efficient for real-world applications.

- *Advantages:*

One of the key advantages of this approach is its **streamlined workflow**, which integrates both text detection and recognition, reducing errors and improving effi-ciency. The **multi-scale feature extraction** ensures that text is accurately recognized regardless of variations in size and orientation, enhancing overall accuracy. The **self-attention mechanism in transformers** allows for better handling of text within complex backgrounds, making it more robust for real-world applications such as street signs, license plates, and store-fronts. Additionally, the framework's **generalization capa-bility** enables it to perform well across different datasets, reducing the need for extensive retraining. These factors make it a significant improvement over conventional OCR mod-els, which often struggle with distorted or low-contrast text. Disadvantages: Despite its benefits, this approach has several challenges. The **high computational cost** associated with transformers makes it difficult to deploy on low-power de-vices, as the model requires substantial processing power and memory. Additionally, the framework **demands large and diverse training datasets** to achieve optimal generalization, which can be resource-intensive and time-consuming to obtain. The **complexity of implementation** is another drawback, as integrating multi-scale feature extraction with transformer-based architectures requires specialized knowledge and longer development times. Furthermore, while the method improves recognition in challenging environments, **it may still struggle with extreme vari- ations in text quality**, such as heavily blurred or occluded text, requiring further refinements for practical deployment.

- *AI-Infused Cloud Storage Management: Struc- tured, Se-cure, Interactive, and Efficient Retrieval System*

This paper explores AI-infused cloud storage management, focusing on secure, efficient, and structured document man-agement. It emphasizes standardized naming, encryption, and user-friendly retrieval strategies.[5] Methodology: The paper "AI-Infused Cloud Storage Management: Structured, Secure, Interactive, and Efficient Retrieval System" presents a cloud-based document management system that leverages artifi-cial intelligence (AI) for automated document segregation, categorization, and retrieval. The system employs machine learning algorithms** to analyze document content, identify key features, and classify files into structured categories for better organization. Additionally, it integrates natural lan- guage processing (NLP)-based AI chat models, allowing users to retrieve documents through conversational queries instead of relying on com-plex search commands. This methodology enhances efficiency by automating tedious manual processes and making informa-tion retrieval more interactive and user-friendly. Advantages: The AI-driven cloud storage system offers several key benefits, including **increased efficiency**, as it automates document categorization and retrieval, significantly reducing manual effort and search time. **Improved accuracy** is another advantage, as AI algorithms consistently apply classification rules, mini-mizing human errors. The system is also **highly scalable**, making it suitable for businesses handling large volumes of digital records. Addi-tionally, the **interactive AI chat

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

feature** enables intuitive and user-friendly document retrieval, eliminating the need for extensive knowledge of search commands, thus improving accessibility for all users. Disadvantages: Despite its advan-tages, the system has some challenges, such as **complex initial setup**, which requires technical expertise to train AI models and integrate them into cloud storage. **Data privacy and security concerns** arise, as storing sensitive information in the cloud necessitates strong encryption and compliance with regulations. The system's effectiveness also depends on **AI model accuracy**, and misclassifications could lead to retrieval errors and inefficiencies. Furthermore, **high computational requirements** for AI-driven processes may increase operational costs, making it less feasible for smaller organizations with limited resources.

III. METHODS AND MATERIALS

The flow of managing the documents is depicted in the figure below.

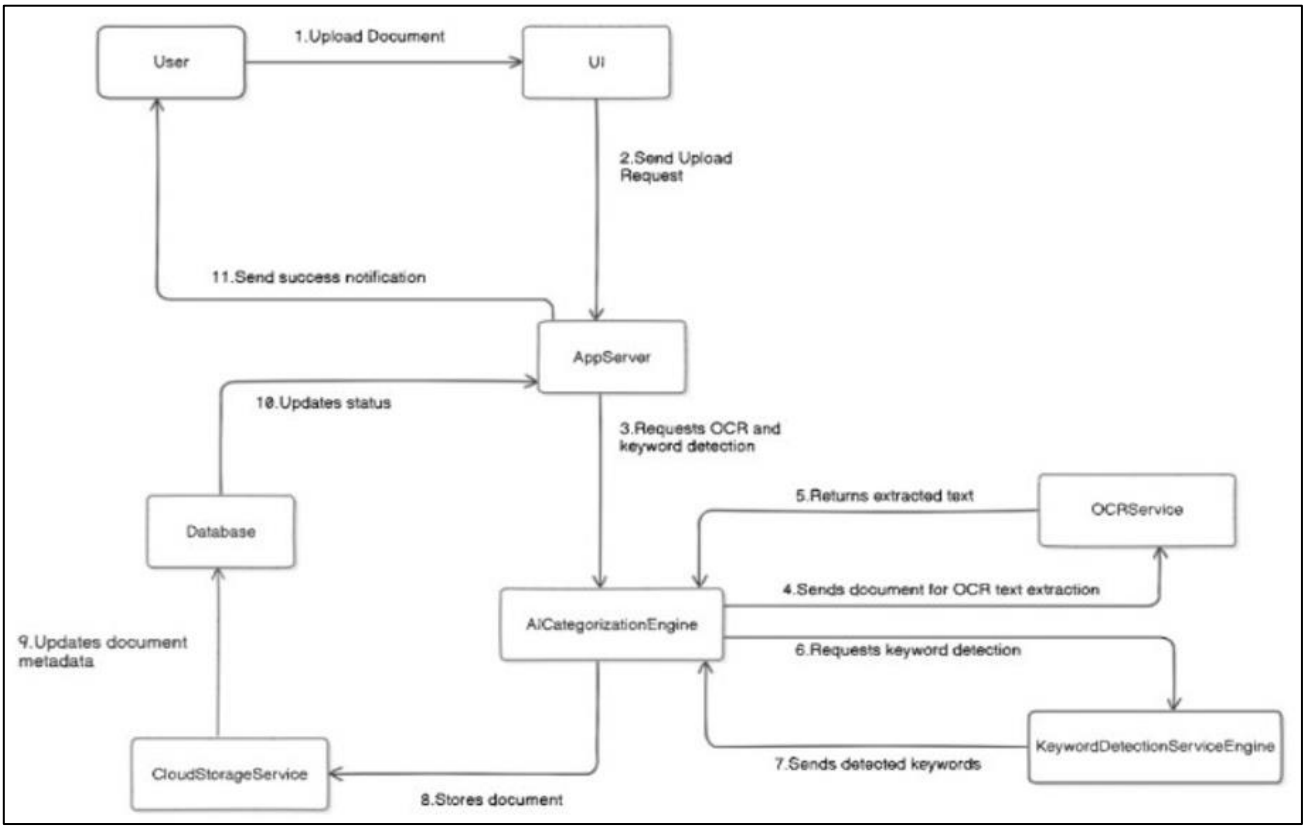

Fig 1 Flow Chart of the Proposed Framework

➢ *Dataset*

The dataset used in the SoftLocker system consists of a collection of digital documents gathered to train, test, and evaluate the AI-based document categorization module. The dataset includes multiple document formats such as PDF files, scanned images (JPEG/PNG), and text documents that commonly appear in organizational and personal document management systems. These documents represent various cate-gories including invoices, receipts, reports, identification docu-ments, application forms, letters, and certificates. The diversity in document types ensures that the system can effectively recognize and classify different real-world documents.

Before using the dataset for training, the documents undergo a preprocessing stage. In this stage, scanned or image-based documents are processed using Optical Character Recognition (OCR) technology to extract machine-readable text. The ex-tracted text is then cleaned by removing unnecessary symbols, correcting formatting issues, and standardizing the content to improve the accuracy of the machine learning model. Each document in the dataset is also labeled according to its category, which helps the AI categorization engine learn patterns, keywords, and contextual relationships associated with different document types.

The dataset is divided into training and testing subsets to evaluate the performance of the AI categorization system. The training dataset is used to teach the model how to identify document categories based on textual features, while the testing dataset is used to measure the accuracy and reliability of the model in classifying unseen documents. By using a diverse and well-labeled dataset, the SoftLocker system can effectively automate document organization, reduce manual classification efforts, and improve the overall efficiency of document management.

➢ *Data Preprocessing*

Data preprocessing is an important step in preparing the dataset before it is used for training the AI-based document categorization module in the SoftLocker system. Since the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

dataset contains different document formats such as PDF files, scanned images (JPEG/PNG), and text documents, the first step involves converting all documents into a machine-readable format. For scanned or image-based documents, Optical Character Recognition (OCR) technology is applied to extract textual content from the images. This process allows the system to read and process information that was originally stored as visual data.

After extracting the text, the data is cleaned and stan-dardized to improve the quality of the dataset. This includes removing unnecessary symbols, correcting formatting issues, eliminating noise, and normalizing the text to maintain consis-tency across all documents. The cleaned text is then organized and labeled according to its corresponding document category, such as invoices, receipts, reports, or certificates. These labeled and structured data are then divided into training and testing datasets, ensuring that the machine learning model can learn patterns effectively and accurately classify new documents during evaluation.

IV. PROPOSED WORK

The proposed work focuses on developing SoftLocker, an intelligent document management system designed to auto-mate the process of organizing, storing, and accessing digital documents efficiently and securely. The system integrates advanced technologies such as Optical Character Recognition (OCR), Artificial Intelligence (AI), and cloud storage to sim-plify document handling. When a document is uploaded or scanned, the OCR module extracts the text from the document, which is then analyzed by an AI-based categorization engine. This engine identifies important keywords and contextual in-formation to automatically classify the document into the most appropriate category, reducing the need for manual sorting and minimizing human error.

In addition to automated categorization, the proposed sys-tem incorporates secure cloud storage using Firebase to ensure safe and reliable document storage. All uploaded files are encrypted before being stored, providing strong protection against unauthorized access. The system also implements role-based access control and user authentication to ensure that only authorized users can view, upload, download, or manage specific documents. Through a user-friendly interface, users can easily upload, search, view, and retrieve documents from any device connected to the internet.

Overall, the proposed work aims to improve the efficiency, security, and accessibility of document management. By com-bining AI-driven automation with secure cloud infrastructure, SoftLocker reduces manual workload, enhances document organization, and provides a scalable solution suitable for both individuals and organizations. This system ultimately enables users to manage large volumes of documents in a structured and efficient manner while maintaining high levels of security and accessibility.

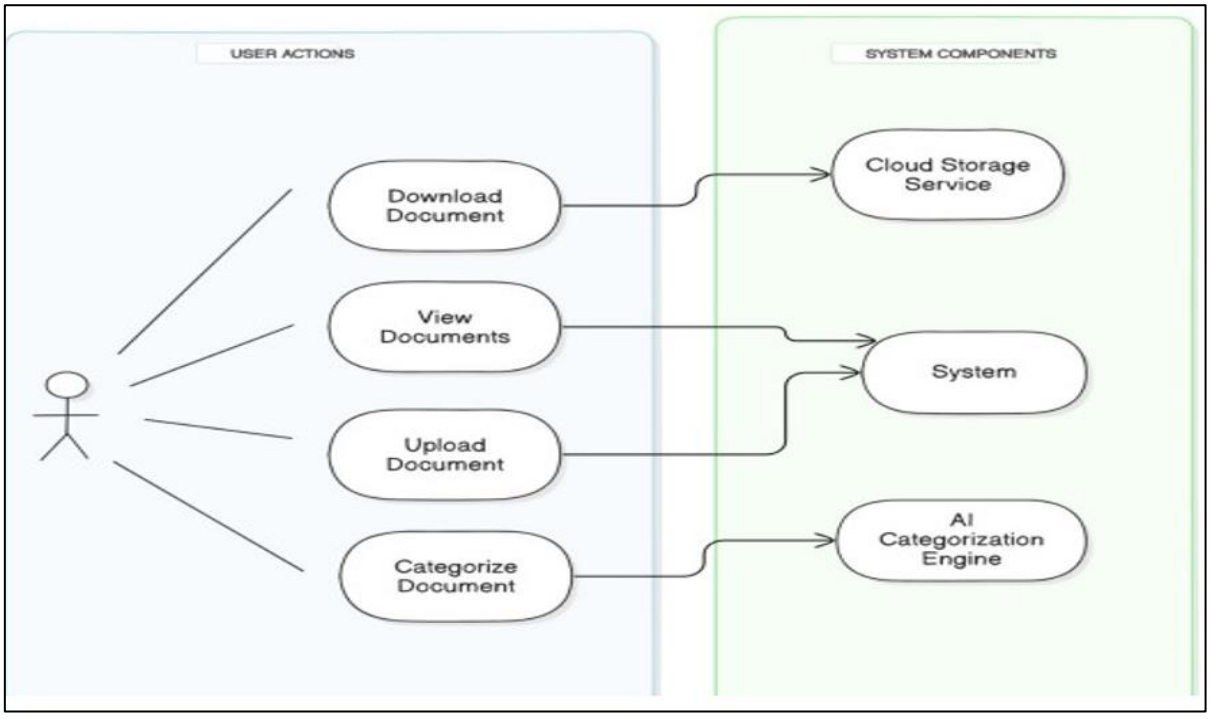

Fig 2 Cse Case Diagrum of Proposed System

Storage ensures secure and remote access, addressing limita-tions in existing document management systems. With a user-friendly interface and robust encryption measures,SoftLocker offers a practical solution for streamlined, organized, and safe document handling, making it highly adaptable for varied user needs in both personal and professional environments. Looking ahead, SoftLocker has the potential to incorporate additional functionalities, such as improved natural language processing for enhanced document search capabilities and advanced machine learning

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may232

algorithms to optimize document categorization. Further integration with cloud platforms and services could expand its scalability and interoperability across systems, potentially leading to seamless connectivity with other business tools and workflows. Additionally, implement-ing continuous learning models would allow SoftLocker to adapt to new document types and user preferences, en-hancing its accuracy and flexibility. With ongoing innova-tion,SoftLocker can contribute to the future of automated document management, pushing industries closer to efficient, paperless workflows. [2] [3]

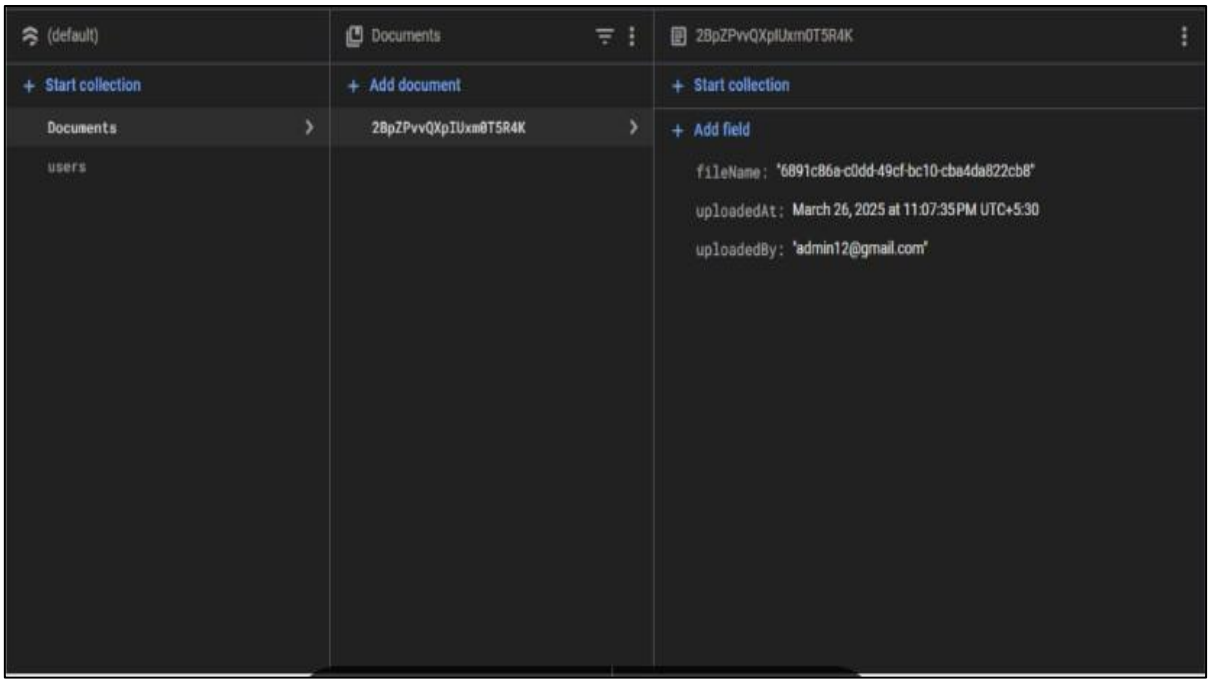

Fig 3 Dcuments Collection in the Firestore Database

Fig 4 LOGIN PAGE

V. CONCLUSION AND FUTURE WORKS

The SoftLocker project provides a significant advancement in document management through automated classification and cloud integration. By leveraging AI and machine learn-ing, SoftLocker efficiently categorizes, organizes, and secures digital and scanned documents, reducing manual effort and enhancing accessibility. The platform's use of Optical Char-acter Recognition (OCR) for text extraction and cloud-based.

REFERENCES

[1]. D. Agarwal, J. Jeevan, R. K. Manikandan, N. Ramith, and M. Vandana, "Advanced automated document processing using optical character recog-nition (ocr)," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1–5, IEEE, 2024.

[2]. P. Bhambri, S. Rani, and P. K. Pareek, "Paperless paradigm: Intelligent automation in document and record management," in *Hyperautomation in Business and Society*, pp. 164–183, IGI Global Scientific Publishing, 2024.

[3]. T. Geng, "Transforming scene text detection and recognition: a multi-scale end-to-end approach with transformer framework," *IEEE Access*, vol. 12, pp. 40582–40596, 2024.