

VAULTX

Akash Rajendran

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

akashrajendran835@gmail.com

Christo V Rajesh

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

christorajesh0@gmail.com

Harikrishnan P T

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

harikrishnan135790@gmail.com

Jebin Joseph

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

jebinjoseph8421@gmail.com

Sam Sebastian

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

samsebastian916@gmail.com

Tinimol Andrews

Department of CSE

College of Engineering Kidangoor

Kottayam, Kerala, India

tinimolandrews@gmail.com

Abstract— The rapid digitalization of financial and administrative processes has led to a significant increase in the volume of documents such as bills, warranties, insurance policies, receipts, and invoices. Recent research has explored the application of Artificial Intelligence (AI), Machine Learning (ML), Optical Character Recognition (OCR), and Natural Language Processing (NLP) to automate document processing, information extraction, and decision support in financial and industrial domains. Existing studies investigate predictive modeling for warranty analysis, AI-assisted warranty claim handling, intelligent document management systems, deep learning-based financial document classification, and OCR-based receipt digitization. Other works focus on secure cloud storage mechanisms, encryption techniques, privacy and ethical challenges associated with AI powered financial systems. This literature review presents a comprehensive analysis of ten representative studies, highlighting their methodologies, key contributions, and limitations. The review reveals that while significant progress has been made in individual areas such as document extraction, analytics, security, and privacy, existing solutions lack a unified, user-centric approach that integrates automated multi-source document ingestion, intelligent extraction, predictive reminders, and secure storage within a single framework.

I. INTRODUCTION

The growing adoption of digital services in finance, insurance, and consumer markets has resulted in the widespread use of electronic documents such as invoices, receipts, warranty certificates, and insurance policies. Managing these documents manually is time-consuming, error-prone, and inefficient, prompting researchers to explore intelligent automation techniques. Artificial Intelligence has emerged as a key enabler for addressing these challenges by facilitating automated document classification, information extraction, predictive analytics, and decision support.

Several studies have focused on warranty and insurance analytics using AI-driven predictive models. Research on warranty notification data demonstrates the effectiveness of machine learning techniques such as logistic regression and time-series analysis in identifying claim patterns and improving risk management. Other studies apply association-rule learning to assist warranty claim handling by recommending related components based on historical data.

In parallel, industry-focused reports highlight the role of AI in insurance ecosystems for claims processing, fraud detection, underwriting, and compliance, while also emphasizing governance and explainability.

Document management and information extraction form another major research direction. AI-based Intelligent Document Management Systems employ OCR and NLP techniques to extract structured information from invoices, identity documents, and financial records. Deep learning pipelines using convolutional neural networks and transformer-based models have achieved high accuracy in classifying and extracting information from financial documents such as receipts and bank statements. OCR-specific studies further address challenges related to noisy receipt images, varying layouts, and multilingual data.

In addition to automation, security and privacy remain critical concerns. Research on encrypted cloud storage proposes end-to-end and end-to-same-end encryption frameworks to protect sensitive user data in outsourced environments. Furthermore, conceptual studies on AI-powered finance discuss data privacy, security risks, and ethical considerations, highlighting the importance of transparency, fairness, and regulatory compliance. Although these studies demonstrate notable advancements across individual domains, they are often limited in scope, focusing on isolated components such as extraction, analytics, or security. This highlights the need to systematically analyze existing research to identify gaps and limitations in current approaches.

II. LITERATURE SURVEY

A. Predictive Modelling and AI Integration for Enhanced Analysis of Warranty and Notification Data

In [1], Siiri Ojala (2024) presented a study on predictive modelling and AI integration for analyzing warranty and notification data within a manufacturing environment. The study focuses on combining warranty claim data and notification records, which were traditionally analyzed separately, into a unified analytical framework. Logistic regression was employed to predict the probability of warranty claims, while time-series analysis was used to identify trends and forecast future claim behavior. A Power

BI-based analytical dashboard was developed to visualize insights and support data-driven decision-making using AI-powered analytics tools.

The results demonstrated a significant positive relationship between the number of notifications and the likelihood of a warranty claim, highlighting the effectiveness of predictive analytics in warranty management. The proposed system enabled improved visibility into claim patterns and supported proactive quality and risk management decisions. However, the study is limited to a single manufacturing case company and lacks real-time integration of predictive models within the analytical dashboard. Moreover, the approach focuses primarily on industrial warranty analysis and does not address automated document extraction from multiple sources such as SMS, emails, or receipts, limiting its applicability to personal financial document management systems.

B. Artificial Intelligence as Assistance in Warranty Claim Handling

In [2], Sebastian Kvist (2023) investigated the use of artificial intelligence to assist warranty claim handling in an industrial setting. The study focused on improving the efficiency of warranty experts by automatically recommending related subparts that are likely to be required along with a primary spare part during claim processing. To achieve this, the work employed an unsupervised machine learning technique known as association rule learning, commonly used in market basket analysis, to identify frequently co-occurring spare parts from historical warranty and notification data.

The proposed approach utilized algorithms such as Apriori to generate association rules based on support, confidence, and lift metrics, enabling the system to suggest relevant subparts during warranty handling. The experimental evaluation showed that appropriate tuning of confidence and support parameters significantly improved recommendation accuracy while reducing false positives. However, the study is limited to a single industrial organization and relies on historical structured data without real-time deployment. Additionally, the system does not address automated extraction of warranty information from unstructured sources such as emails, SMS, or receipts, which restricts its applicability for personal-level financial document organization systems.

C. AI for Better Insurance: Enhancing Customer Outcomes Amid Industry Challenges

In [3], Bratanova et al. (2025) presented a comprehensive industry-focused report on the adoption of artificial intelligence in the general insurance sector to improve customer outcomes and operational efficiency. The study was conducted by CSIRO in collaboration with the Insurance Council of Australia and examined how AI technologies are being applied across the insurance value chain, including automated claims processing, fraud detection, underwriting, risk assessment, and compliance monitoring. The report highlights the role of machine learning, natural language processing, optical character recognition, and predictive analytics in reducing manual workloads, improving accuracy, and enabling faster customer service.

The findings indicate that AI-driven document processing and predictive analytics can significantly enhance claims

triage, policy management, and customer experience by automating information extraction and decision support. The study also emphasizes the importance of governance, transparency, and explainability in AI systems to maintain consumer trust, particularly when handling sensitive financial and insurance data. However, the report is largely conceptual and policy oriented, with limited focus on concrete system implementation or user-level applications. Additionally, it does not address personal document organization across multiple data sources such as SMS, emails, and receipts, nor does it provide a unified secure platform for managing warranties, payments, and insurance documents, which limits its direct applicability to individual financial document management systems.

D. AI-based Integrated Approach for Intelligent Document Management System

In [4], Pandey et al. (2023) proposed an AI based Integrated Intelligent Document Management System (IDMS) to automate the extraction and processing of information from structured and unstructured documents. The system leverages Artificial Intelligence, Machine Learning, Natural Language Processing, and Optical Character Recognition techniques to extract key information from medical invoices, Aadhar cards, and PAN cards. A hybrid approach combining Computer Vision-based OCR and NLP using regular expressions was employed to improve extraction accuracy and robustness across different document formats.

The proposed IDMS was evaluated using datasets consisting of hospital invoices, Aadhar cards, and PAN cards, and its performance was compared against EasyOCR. Experimental results demonstrated that the hybrid CV+NLP approach achieved superior accuracy, reporting 97% accuracy for medical invoices, 71.87% for Aadhar cards, and 78.95% for PAN cards. Although the system effectively automates document digitization and structured data extraction, it is limited to a predefined set of document types and operates primarily on scanned images. Furthermore, the study does not incorporate automated ingestion from communication channels such as SMS or emails, nor does it provide predictive reminders or renewal alerts, which are essential features for personal financial document organization systems.

E. Deep Learning for Effective Classification and Information Extraction of Financial Documents

In [5], Serbanescu and Dhali (2025) proposed a deep learning-based pipeline for the classification and information extraction of financial documents such as invoices, receipts, bank statements, and payment records. The study addresses the challenges of large-scale financial document processing by integrating Optical Character Recognition with advanced deep learning models. For document classification, several convolutional neural network architectures including LeNet5, VGG19, and MobileNetV2 were evaluated, while information extraction was performed using transformer based and multimodal models such as RoBERTa, LayoutLMv3, and GraphDoc.

The experimental results demonstrated that MobileNetV2 achieved superior performance in document classification, with an accuracy of 99.24% when data augmentation was

applied, whereas RoBERTa and LayoutLMv3 produced the highest F1-scores in information extraction tasks. The study highlights the effectiveness of combining visual, textual, and layout-aware representations for automating financial document processing. However, the proposed pipeline is evaluated mainly on scanned and photographed documents and relies on offline batch processing. Furthermore, it does not support automatic ingestion of documents from communication sources such as emails or SMS, nor does it provide predictive reminders, renewal tracking, or secure personal document organization, which limits its direct applicability to user-centric financial management systems.

F. Automating Information Extraction from Textual Documents Using NLP Techniques

In [6], Somaya et al. (2024) presented a comprehensive review of Natural Language Processing (NLP) techniques for automating information extraction in Document Management Systems (DMS). The study examined the evolution of NLP approaches from traditional rule-based methods to modern deep learning models, with a focus on key techniques such as Named Entity Recognition (NER), syntactic and dependency parsing, and sentiment analysis. These techniques form the foundation for extracting structured information from unstructured and semi-structured textual documents.

The review highlighted how integrating NLP into DMS enables automated document indexing, categorization, metadata generation, and improved retrieval efficiency across multiple domains including healthcare, finance, legal services, and human resources. The authors also discussed real-world challenges such as language ambiguity, data privacy concerns, and scalability issues when deploying NLP-driven systems. However, being a survey-based study, the paper does not propose a concrete system architecture or implementation framework. Additionally, it does not address multi-source document ingestion from channels such as SMS or emails, nor does it incorporate predictive features like expiry tracking or renewal reminders, which are essential for user-centric financial document organization systems.

G. BillBuddy: AI-Powered Receipt Capture and Smart Bill Sharing for Expense Management

In [7], Suhaimi et al. (2025) introduced Bill-Buddy, an AI-powered mobile application designed to simplify personal and group expense management through automated receipt capture and processing. The system employs Optical Character Recognition (OCR) and Artificial Intelligence techniques to scan receipts and automatically extract, categorize, and store expense-related information, thereby reducing manual data entry and improving accuracy. In addition to individual expense tracking, the application supports shared expense management by enabling users to create group events, split bills equitably or proportionally, and track individual contributions.

The study demonstrated that integrating AI-based receipt scanning with smart reminders enhances transparency and accountability in shared financial activities. Automated payment reminders and contribution tracking help users avoid missed payments and reduce social friction during collaborative spending scenarios. However, the proposed system is primarily focused on expense tracking and bill

sharing and does not support long-term document organization such as warranty or insurance management. Furthermore, it does not incorporate secure archival mechanisms for sensitive financial documents or predictive features for expiry and renewal tracking, limiting its applicability to comprehensive personal financial document management systems.

H. End-to-Same-End Encryption for Secure Cloud-Based Data Storage

In [8], Chen et al. (2025) proposed an end-to-same-end encryption framework that enables secure outsourced data storage on existing cloud infrastructures without requiring modifications on the cloud server side. The study addresses the challenge of protecting user data from cloud service providers by ensuring that only legitimate users can access their stored data across multiple devices using a single passphrase. The proposed system builds upon existing encryption mechanisms and introduces a modular architecture that supports portability, efficiency, and user mobility while maintaining strong privacy guarantees.

Experimental evaluation conducted on Amazon S3 demonstrated that the proposed encryption scheme achieves practical performance with minimal overhead while ensuring confidentiality against threats from cloud servers, key servers, and unauthorized users. The system also supports seamless integration with existing applications and login mechanisms, making it suitable for largescale deployment. However, the work primarily focuses on secure cloud storage and does not incorporate document-level intelligence such as automated information extraction, classification or predictive analytics. Additionally, it does not address application-specific features such as reminders, expiry tracking, or personal financial document organization, which are essential for user-centric secure document management systems.

I. Deep Learning-Based OCR Engine for Printed Receipt Images

In [9], Sayallar et al. (2023) proposed a deep learning-based Optical Character Recognition (OCR) engine, named Nacsoft OCR, specifically designed for printed receipt images. The study addresses the challenges of digitizing receipts and invoices that often suffer from noise, background clutter, perspective distortion, varying fonts, and poor image quality. The proposed OCR pipeline consists of three main stages: preliminary image preparation, word position detection, and word reading. Deep learning techniques such as convolutional encoder-decoder networks for receipt segmentation, YOLO-based object detection for word localization, and Convolutional Recurrent Neural Networks (C-RNN) for text recognition were employed.

The proposed OCR engine was evaluated against widely used OCR systems including Tesseract, EasyOCR, and Google Vision API using both Turkish and English receipt datasets. Experimental results demonstrated that the Nacsoft OCR engine achieved superior performance in character recognition accuracy and word position detection on Turkish receipt data, while showing competitive performance on English receipts. However, the system is trained primarily on receipt images and is limited to offline image-based processing. Additionally, it does not incorporate higher-level

semantic information extraction, automated ingestion from sources such as emails or SMS, or secure long-term document storage with predictive reminders, which are essential features for comprehensive personal financial document management systems.

J. Data Privacy, Security and Ethical Considerations in AI-Powered Finance

In [10], Brightwood and Jame (2024) discussed the critical issues of data privacy, security and ethical considerations arising from the increasing adoption of artificial intelligence in financial applications. The study highlights how AI-powered financial systems improve efficiency, risk management and personalized services, while simultaneously introducing significant risks related to the handling of sensitive personal and financial data. The authors emphasize the importance of safeguarding user data against unauthorized access, breaches, and misuse, particularly as large volumes of customer information are collected and analyzed.

The paper also examines ethical challenges such as transparency, accountability, and trust in AI driven financial systems, stressing the need for robust privacy-preserving mechanisms and secure data management practices. While the study provides valuable conceptual insights into privacy and ethical risks, it does not propose a concrete technical architecture or implementation strategy. Furthermore, it does not address application-level solutions such as encrypted personal document storage, automated document extraction, or user centric reminder systems, which limits its direct applicability to practical AI-based personal financial document management platforms.

III. CONCLUSION

This literature review examined ten research works addressing various aspects of AI-driven document processing, financial analytics, and secure data management. The reviewed studies demonstrate that AI, ML, OCR, and NLP techniques are highly effective in automating document classification, extracting structured information, and supporting decision-making in warranty management, insurance, and financial applications. Deep learning models such as CNNs, transformers, and multimodal architectures have shown strong performance in handling complex document layouts and unstructured data. Additionally, encryption-based cloud storage solutions and discussions on data privacy and ethics underline the importance of safeguarding sensitive financial information. Despite these

advancements, the review identifies several research gaps. Most existing systems operate in isolation, addressing only specific tasks such as warranty prediction, receipt OCR, or secure storage. Few studies consider automated ingestion from multiple real-world data sources such as SMS, emails, and scanned receipts simultaneously. Moreover, predictive reminder mechanisms for expiry dates and renewals are largely absent, and security solutions are often treated separately from intelligent document processing. Ethical and privacy-focused studies, while informative, lack concrete technical integration with document management systems. Overall, the literature indicates a clear need for a unified, intelligent, and secure document management framework that combines automated multi-source data acquisition, AI-based information extraction, predictive analytics, and robust security mechanisms. Addressing these gaps can significantly enhance efficiency, reliability, and user trust in future AI-powered financial document management solutions.

REFERENCES

- [1] A. G. Gabriel, A. A. Ahmad, and S. K. Jeyakumar, "Advancing agentic systems: Dynamic task decomposition, tool integration and evaluation using novel metrics and dataset," arXiv preprint arXiv:2410.22457, 2024.
- [2] D. B. Acharya, K. Kuppan, and B. Divya, "Agentic ai: Autonomous intelligence for complex goals—a comprehensive survey," IEEE Access, vol. 13, pp. 18912–18936, 2025.
- [3] L. MaChucho and J. Ortiz, "The impacts of artificial intelligence on business innovation: A comprehensive review of applications, organizational challenges, and ethical considerations," Unpublished Manuscript, 2025.
- [4] S. Kumar and R. Singh, "Chatgpt and other ai copilots as knowledge partners for entrepreneurs and small business owners: A study of ai copilots' adoption as a medium of entrepreneurial training," Scholedge International Journal of Management & Development, vol. 10, no. 1, 2023.
- [5] H. Dhiman, C. Wächter, M. Fellmann, and C. Röcker, "Intelligent assistants: Conceptual dimensions, contextual model, and design trends," Business & Information Systems Engineering, vol. 64, no. 5, pp. 645–665, 2022.
- [6] S. Mishra and A. R. Tripathi, "Ai business model: An integrative business approach," Journal of Innovation and Entrepreneurship, vol. 10, no. 1, p. 18, 2021.
- [7] M. Mekni, "An artificial intelligence-based virtual assistant using conversational agents," Journal of Software Engineering and Applications, vol. 14, no. 9, pp. 455–473, 2021.
- [8] J. T. Cao, Y. Q. Khoo, S. Birajdar, Z. Gong, C.-F. Chung, Y. Moghaddam, A. Xu, H. Mehta, A. Shukla, Z. Wang et al., "Designing towards productivity: A centralized ai assistant concept for work," The Human Side of Service Engineering, p. 118, 2024.
- [9] F. Ferrati, P. Kim, and M. Muffatto, "Generative ai in entrepreneurship research: Principles and practical guidance for intelligence augmentation," Foundations and Trends® in Entrepreneurship, 2024.
- [10] Y. Zhang and K. Siau, "Meta-entrepreneurship: An analysis theory on integrating generative ai, agentic ai, and metaverse for entrepreneurship," Journal of Global Information Management, vol. 32, no. 1, pp. 1–21, 2024