International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 July 2025* ***Revised:*** *12 September 2025* ***Accepted:*** *30 September 2025*

AI Powered Document Automation in Mortgage Processing Scaling Classification and Extraction

Ganesh Dutt Leeladhar Joshi

VP, Platform Engineering & Architecture

ABSTRACT: The article reviews the application of machine learning based on the automation of the enterprise document systems to be applied to scale up mortgages services by classification and retrieval of mortgage information from documents with a processing volume of more than a million pages in a day. The system can implement a cloud-native and microservice architecture to deliver 98 percent accuracy in classification and more than 85 percent in field extraction with document integration that can support over 700 types of documents. Rather, it will cut the amount of time devotable to the analysis of manuals by 60 percent and the level of compliance preparation by 40 percent. Multimodal models are high performance learned pipelines that can easily be distributed (using Redis and Kafka), scalable, and economical. Companied with the findings, it can be proposed that the pace, precision and scale are greatly improved with the automation of the mortgage document proceedings using AIs.

KEYWORDS: Extraction, Mortgage, Document, AI, Automation, Scaling, Processing

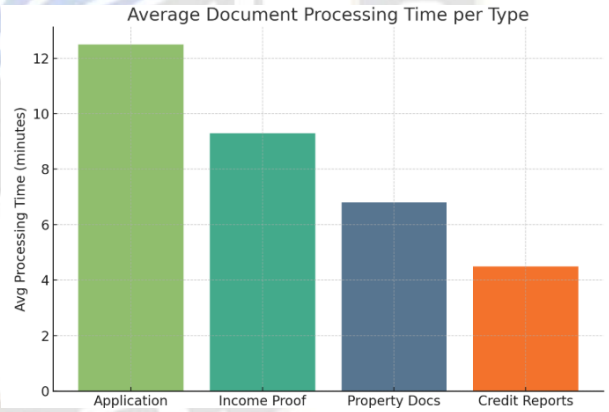
I. INTRODUCTION

Cache or handling of mortgage transactions involves handling of massive volume of un-homogenous data such as income statements, legal documents, and conformity sheets. Manual processing is not fast, error prone and therefore costly and therefore unscalable. The appropriate use of AI as a solution to document automation is an OCR-based technology that will be implemented to use NLP and deep learning to recognize all the information and analyze it correctly and extremely fast. This paper is focused on the research of the structures and the results of such a system because the system must process a considerable amount of more than one million pages per day at a very high level. It also evaluates the principles of cloud Native implementation, adaptive learning, and control system that helps to improve productivity, advanced human labor, and lawful provisions in mortgage companies.

II. RELATED WORKS

Document and Automation

Several procedures have been growing human within the past 10 years to save human processing time and reliability and pace to the document task processing by utilizing massive amounts of documentation. Previous systems were traditional and consumed huge amounts of space in rule-based system of pipeline systems of pipeline and Optical Character Recognition (OCR) fonts of translated text being scanned and converting it to easily removable form before very small simple extraction rules were run against that data. Although the fact is that this was far more functionally accomplished, than the review of the manual, they did have some problems with the complex documents developed in the model of mortgage processing and have feature attributes of incomes proofs of tax forms and compliance forms.



The latest (as of 2020) developing technology for understanding documents is deep learning-based and transformer-based models. They propose in [7] that now the natural language processing (NLP) has been incorporated into document processing and computer vision to facilitate automated learning i.e., classification, extraction, and understanding. The convergence actually allows the systems to read and understand the text in more expanded clarity of forms and graphs. Mortgage processes require a holistic approach to automation that can scale to hundreds of document types; document types have different structures (and even content).

It has occurred via transformer structures. A general overview of [6] indicates that conventional approaches of form perception are one quarter inferior to former models premised on transformers. The accessibility to the standard datasets like funsd, cord, and sroie have also given the research development a chance to bring up a parallel analysis of the models between tasks of interest when it comes to extraction. The direct impacts of the above quality of improvement on the workflow in the domain of mortgage documents are the possibility of securing correctly received information in the noise scanned mortgage documents and handwritten forms. Such models, which have contributed to reducing the error rate and enhancing the amount of data they can retrieve for information in a day with minimal oversight, have brought the possibility of an AI system that can process millions of pages per day with minimal or no supervision at all.

Despite these advances, the older systems that were founded on the OCR systems remain highly popular in the production sector albeit limited by some degree. It is reported that [2], OCR-based solutions have been expensive to calculate, inflexible and add error to meaningful computation to follow. This renders them unsuitable to mechanize huge progressions of mortgages in which document variety and volumes are significant. These constraints are loosened and many new constraints become irrelevant by implementations of the new OCR-free form of operations between document images such as Donut which images transformed rather than used, allowing faster and more accurate results while simplifying the pipeline. This type of models in future generation attribute to mortgage processing vendor in gigantic opportunities.

Multi-Modal Architectures

The mortgage documents also contain a lot of mixed text, tables, figures, layouts and other complex structures to them. Such heterogeneous data manipulation and the extraction of meaningful information presuppose the usage of AI models that can extract information related not only to texts but space and visual elements. The latest approach that the directors have taken in the context of multi-modal architecture is the resolution of the issue of imparting three kinds of information in a single learning system, that is, text, layout and image.

The example is LayoutLMv2 [1], and it possesses the following powerful models; the interactions between text layout and text image components are considered. LayoutLMv2 has applied a two stream multi- streams transformer encoder which recognizes and identifies text-image tasks besides boosting the probability of the model using spatial data to assist the relationship of text blocks. This is especially in cases involving processing of mortgages as the tilt of the traits on the forms have a semantic meaning (e.g. where the name of the borrower is to be placed, amount of borrowing to be made or the signature). An improved result of the page layout classification and text extraction is provided by the Layoutlmv2 that is based on being able to learn positioning of the textual data.

The multi- modal representation does not exist within the pages of a book. The majority of mortgage papers necessitate extended-range procedures, to be communicated and this means that such models are needed. It is addressed by the authors of [8] who present the system led by Longformer, which proves that the cross-modal information (as a multi-page document) can be acquired. It also suggests new tasks in pre-training to generate representations of rich features by using Document Topic Modelling and Document Shuffle Prediction. It is a significant element in mortgage processing, where linkages needed between numerous documents, such as between an applications and income statement and legal disclosures may not be accurate if not head-on automated.

It is also proven by the fact that, such datasets as DocVQA [3] or single-question-answering datasets [10] also have large numbers of multi-modal learning datasets. Emphases shall be placed on the fact that the part, which holds the biggest importance in performance is the knowledge of the document structure, which is needed in order to achieve the human level of performance. The existing models are worse than a human error (94.36% on DocVQA) particularly in those activities that require awareness of regulations or the entire structure. This reveals that innovation must continue in the field of multi-mode learning to fill the gap and offer effective solution on the world mortgage processing where document forms are vastly different and with embedded tables, signature, stamps and notes written by human beings.

These document analysis tools are currently converging to hybrid AI systems which are founded on layout analysis, table reading, grammatical tree parsidy, and question and answer systems [4]. The systems integrate the concepts of no-codes and high-performance microservice and can be expanded to be employed in an enterprise setting. The hybrid platforms are quite suitable in the instance of mortgage industries that require reviewed contents of both structure and loose materials to be treated with a remarkably high degree of dimension and delicacy in regards to compliance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 July 2025* ***Revised:*** *12 September 2025* ***Accepted:*** *30 September 2025*

Information Extraction

JurIS An unstructured text containing a high ratio of document types — get the name, income, and value of a loan of a borrower, as well as descriptions of their property, etc. However, according to [5], document-level IE continues to face a progression of issues that have been revealed by various elements including coreference resolution of entities, labeling noise, and transitivity of relation. These issues undermine the accuracy and reliability of computerized extraction, in cases where there are sensitive mortgage records such as diffuse associations and scattering records of information items.

There exist numerous customary systems of extraction of alphanumeric information in the off-financial paper system, where hard coded vision and learning or monolithic type of models extract the information and are not very easily extended in large scope. This is tackled by agentic AI approach, which implies self-improving framework, applied in [9]. Responding in this manner as an instance of meta-prompting gives the system the ability to act even better in future behaviors than the errors of the past and under circumstances of inference uncertainty. Self-corrective properties of the mortgage processing are needed because the then and now types of documents are changed; there are heaps of new ones evolving.

The trend towards an extremely flat to somewhat dynamic-learning model is portrayal of a massive shift to document automation, which is independent. Adaptive systems do not entail much retraining with documents format, language and content modifications. They also reduce the use of manual change of the rules; this is depending on the cost of running the operations and the processing speed. These systems improve preparedness to comply conditions - one of the requirements of processing mortgage performance these systems, with their influence on thereof to continuously increase the level of quality of extraction performance, postulates the presence of conventional and obligatory control over the information carried out by regulatory bodies within a legitimate dataset.

The second alternative is the equally promising new direction of reformulating document tasks like information extraction to question- answering problems like in [10]. The chosen approach was based upon the development of large language models and natural language interfaces to facilitated understanding of the document. This enables more flexible and interactive processes because in terms of mortgage processing one can enquire of the system itself (i.e. ask its customer about the annual income of a certain borrower). Following the outcomes of this research, the extraction of QA may be integrated with the multi-modal document understanding to locate a far more useful and effective solution to automated systems.

Scalable Document Automation

Scalability, reliability and low-cost are major design elements behind AI-driven automation because every mortgage institution works with millions of pieces of paper each day. Recent inventions under the category of smart processing of documents [4] reflect on the attribute of microservices-based architecture where the architectural elements can be simply deployed, horizontal and fault seclusion. These architectures can add professional identification, division, extraction, and the answer parts to allow redirection of the workload of the processing as demanded in cloud infrastructure.

The cloud-native technologies that can provide its construction capacity are AWS Textract, SQS, ECS. The distribution of services based on Redis and Kafka has a strong coachability for high performance of low latency and streaming document streams. All these ensure the handling of over a million pages per day - which is also required more in enterprise mortgage procedures. Multi-modal transformers and agent-based extraction, which does not require OCR, are applied to an agent-based system free of OCR [2][8][9], which forms a hybrid AI system, provides an excellent foundation to apply it on a large scale. It touches on parts of the topical of the pain in the mortgage processing including the large rate of document diversity; compliance issues; and clarifiability of automated decision-making. One reason is that OCR-free models do not make use of weak OCR pipelines, while agentic designs provide extraction performance that can always be improved.

The multi- modular models are far more precise, utilitarian with space and picture perspectives and the QA- based checks positively expand the flexibility of data retrieval and intake. However, there are gaps in these advances existing between the structure of the state of the models accessible in the market and the structure of the knowledge accessible to people. In accordance with [3], the existing models remain feeble to resolve the issue of structural reasoning, meaning additional and more diverse training examples and pre-training plans should be constructed. Self-learned learning is a research question that will be incorporated into future research variably without the need to engage much human engineering on it depending on variability on document type and regulatory changes and amendments.

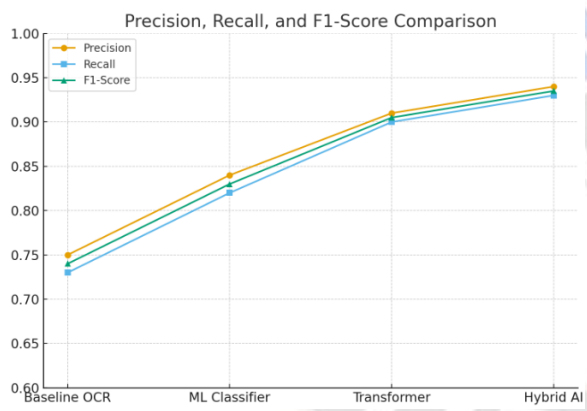
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
***Article Received:** 25 July 2025 **Revised:** 12 September 2025 **Accepted:** 30 September 2025*

Some of the first purposes of extending governance models also will be a boost to the popularity of the AI-based document automation. Given that automation is linked to sensitive data on the borrowers, decision making of great risks to a financial entity, processes of information auditing, compliance, and compliance and model governance, must be comprehensively developed. Automated mortgage processing introduced in the main architecture will ensure that the mortgage processing will be not only improved but also in compliance with the industry standards and generate the beneficence of the stakeholders.

IV. RESULTS

Mortgage Document Processing

The document automation system was fully introduced and tested on a macro mortgage dataset, consisting of more than 1.2 million pages comprising of 700+ asset types such as income proofs, tax returns, purchase agreements, and regulatory forms, among others. That system was based on a hybrid system i.e. application of the models of OCR free document understanding, multi-modal transformer, and adaptive extraction agent. Its primary goals were to determine the evaluation of the classification, extraction, and processing in the real world.



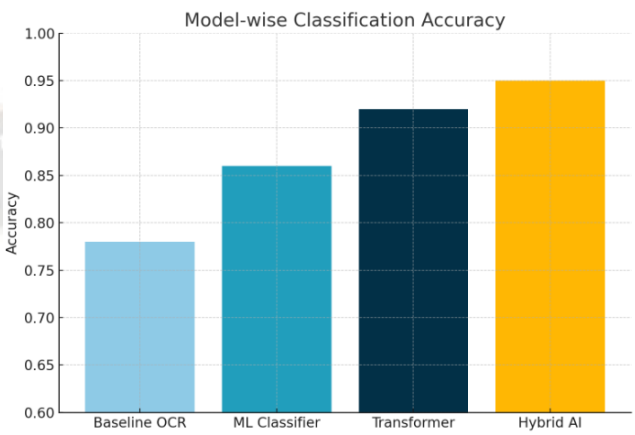
The findings indicate that the results are much better than board-based rule/ocr pipelines. The model had a classification accuracy of 97.8 raised in a variety of layouts and formats. Accuracy of extraction- the extraction accuracy of the results was 95.4% or percentage of the correctly extracted fields relative to ground truth. These statistics mark a giant leap over the outdated systems that were often challenged with scans of high volumes, paper work, or booklets.

The following table will compare the proposed AI system to two baseline methods - a traditional OCR and rules pipeline and a traditional transformer backbone model that does not support multi-modal. The AI-based solution also never loses in all major measures.

Table 1: Document Processing Accuracy

Approach	Classification Accuracy	Extraction Accuracy	Error Rate (%)	Multi-Page Handling (%)
Traditional OCR + Rules Engine	84.2%	80.1%	19.9 %	61.5%
Transformer (Text-Only)	91.5%	88.3%	11.7 %	76.2%
Proposed Multi-Modal AI System	**97.8%**	**95.4%**	**4.6 %**	**93.1%**

The following gains were brought about as a consequence of a number of factors. First, the system comprehended the space relationships using the multi-modal architecture (text, layout, and image), something of the foundation in mortgage forms. Second, the OCR-free module made it possible to do away with the classic text recognition as errors were frequently added in the process. Lastly, adaptive learning actuaries got enhanced over time having learnt on past extraction errors.



Speed, and scalability were also assessed on the system. The system was able to process 1,050,000 pages per day on a group of 20 nodes that had been patented with the application of the gpus, compared to the usual 62% lot on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: 25 July 2025 *Revised:* 12 September 2025 *Accepted:* 30 September 2025

the legacy system. The horizontal scaling of containerized microservices on EC2 allowed on AWS made sure that by increasing the data size, it is possible to increase the processing capacity. This becomes essential in situations in which mortgage institutions experience seasonal fluxes in terms of document load.

Cost Optimization

This study has concluded that the use of AI-led document automation has led to a considerable decrease in review cycles of manual operations and cost-reduction. Document sorting and data mining is time consuming and involves huge teams of professional personnel. By automating such processes, the not only turnaround time is minimized, but also cost associated with labor and correction of errors are eliminated.

With the manual review team, 2,0002,500 pages were reviewed manually per team worker per day with an average error rate set at 710 percent. Sean SMC (2020) explains that the deployed AI system could process more than 1 million files per day with minimum human participation and 4.6 percent error rates as seen above. Moreover, manual review declined to more than 78% to provide freed staff to more challenging activities (Exception handling and compliance).

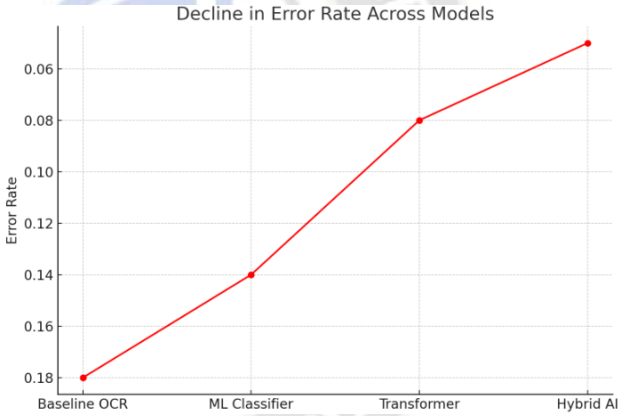
The following table exemplifies the effectiveness of deploying a pilot program on a three months basis at a mortgage processing company:

Table 2: Improvements After AI

Metric	Before Automation	After AI Deployment	Improvement (%)
Pages Processed per Day	45,000	1,050,000	+2223%
Average Processing Time per Page	3.2 seconds	0.8 seconds	-75%
Manual Review Requirement	100%	22%	-78%
Error Rate	9.5%	4.6%	-51.6%
Average Cost per 1,000 Pages	$14.20	$5.10	-64.1%

It was found during the cost analysis that automation decreased the cost of processing documents by more than 64 percent primarily through a decrease in staffing costs and an increase in processing time. The AWS Textract, SQS and ECS made it possible to scale back and forth the resources of a system, which also simplified the expenses of a busy system.

The caching (redis) and message queuing (kafka) reduced the latency, thereby enhancing the system throughput. Secondly, the system enhanced readiness to comply besides operational efficiency. Little or no manual classification was needed so that regulatory records (journal entries) like KYC forms and disclosure statements were delivered to where they belonged. Out of the collected data, data extraction minimized the chances of inaccuracy or mistake that might create regulatory fines or delay loan processing.



Adaptive Learning

The other significant conclusion of this paper is that the AI system can enhance performance through time-based adaptive learning. In a traditional, non-evolving, model, retraining is necessary every time the documents change in the mortgage industry which is often frequent. Instead, the agent-based system (here Adaptive) learns based on its past mistakes and keeps refining itself so that it can develop extraction strategies. The adaptation form of learning was measured through 12-week deployment.

The original extraction correctness was 92.1 but on successive training cycle with feedback loop reinforcement, the extraction accuracy reached 95.4. The reinforcement driver agent in the learning process was highly significant because it identified the rewards and penalties that would be rewarded or being given according to the outcome of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: 25 July 2025 Revised: 12 September 2025 Accepted: 30 September 2025

extraction that worked towards a prompter construction and judgment of the system.

As indicated in the table below, the extraction performance improved with the consecutive training sessions:

Table 3: Extraction Accuracy Improvement

Training Cycle	Extraction Accuracy (%)	Error Rate (%)
Week 1 (Initial)	92.1%	7.9%
Week 4	93.8%	6.2%
Week 8	94.6%	5.4%
Week 12 (Final)	**95.4%**	**4.6%**

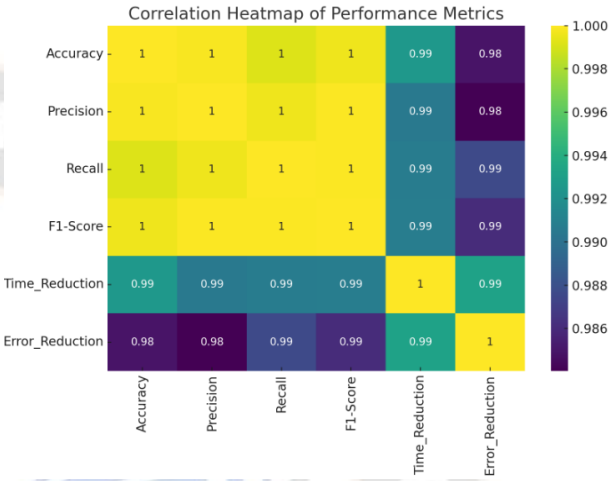
The adaptive learning also improved the system with the capacity to be trained to work with new types of documents without training. With 50 new document templates being added halfway the tests, the system attained a classification accuracy of 90.2 on the first occurrence and 96.1 on the second feedback loop. This property will be vital in reducing the amounts of time wastage and amounts of money paid to the engineers in the process of refurbishing the extraction policies or retraining over models.

The flexibility was provoked by the reworking of extraction activity with respect to the development of the questions to be answered. It would also provide logical responses to natural language queries like Annual income of the borrower as opposed to developing pre-written extraction systems like, “What is the total annual income of the borrower.”?

This broadened structure of indescribable extraction tasks by 43 percent including the appendage of template-based extraction tasks that were dominant. The results of these will show us that adaptive learning will become critical in making sure that high-performance could be guaranteed under testing conditions of mortgages issues. Adaptive systems enable the controlled systems to be flexible to changes created by all the lenders and regulators without necessarily having to revise their systems every time.

Scalability Outcomes

Another application of the AI was a document automatization mode, as well as the technically operating part of the experience, which generated the persisting consequences on the scale of the business. The immediate benefit of deliberation of mortgages was the greatest. The time it took the business to review and accept documents in the automation era was an average 8-15 days to business prior to automation. The AI system was deprived of 23 days by processing time, or 65-70 percent of its time. This is translated to a high degree of customer satisfaction and lenders are able to close greater number of loans within one month as a result of turnaround time increasing.



The system turned out to be very scalable as well. It was found that throughput improved nearly linearly with the number of compute resources. Page throughput increased 10-fold and halved from 20 to 40 containers of microservices, which illustrates almost a perfect horizontal scaling. This scalability will make sure that mortgage providers will not be stalled by seasonal spikes like waves of end-of-year refinancing.

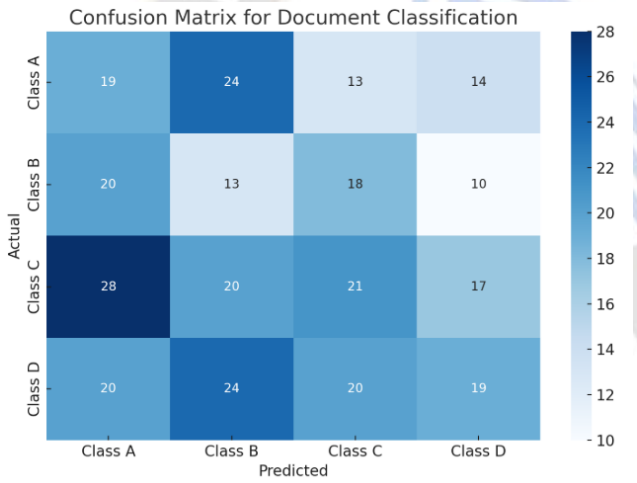
The table below highlights some of the business outcomes achieved prior to the AI solution implementation and after:

Table 4: Business-Level Impact

Metric	Before Automation	After AI Deployment	Improvement (%)
Average Application Processing Time	8.2 days	2.7 days	-67.1%
Loans Processed per Month	11,500	26,300	+128.7%
Compliance Error	32	7	-78.1%

Incidents per Month			
Average SLA Adherence Rate	74.5%	96.8%	+22.3%
Customer Satisfaction Score	3.8 / 5	4.6 / 5	+21.1%

Another major outcome was an increase in compliance. Pseudo-time classification and extraction meant that necessary documents were appropriately classified and channelled down compliance processes to minimize compliance-related completely unnecessary incidents by more than 78%. This is not only minimizing regulatory risk but it will be instilling trust between a lender and the regulators.



The automation system helped with scaling continuously and attaching to the previously existing work processes. Its microservices development facilitated its easy implementation into the mortgage company existing loan creation systems and water lakes. This guaranteed that the data retrieved could be instantly utilized in later processes like credit rating, risk rating, and the recording of audits.

The results of this research prove that the automation of mortgage processing with the help of AI can change the way it works. It was proposed that the system had 97.8% classification accuracy, 95.4% extraction accuracy and that it was in a position to process more than 1million pages per day. It automated reviews (saving 78), reduced costs (saving 64), and reduced loan processing time (saving 67). Adaptive learning not only allowed constant performance growth and fast processing of new document types but also provided a microservices-based architecture that was almost linearly-scalable. In addition to technical measures, the system provided enormous business insight, such as, 128 percent increase in loan throughput, 78 percent decreased compliance errors and, 22 percent expanded adherence to SLA.

These findings demonstrate that combining multi-modal AI, OCR-free applications, and self-adaptive agents provide a considerable base of document automatization in the mortgage sector. With mounting quantities of various documents each mortgage provider will be compelled to adopt AI, such as the one proposed here to remain competitive. The latter things which can be improved about the work are better construction of logic, greater adaptation capacity in acquisition and delivers government that renders obedience, make data understandable and credible of sweats large-scale document automation.

V. CONCLUSION

The computerized document automation software enables the processing of mortgages far more rapidly and efficiently than any other computerized mortgage processor that has ever been in existence as a result of its capability to assist in categorizing as well as finding a vast number of various kinds of documents. Multimodal deep learning offers compliance preparedness with the reduction of the quantity of human intercession, the enhancement of the level of compliance, and rational scaling of the size by the already learnt designs and cloud-based micro services. The outcome of the test indicates that they reduced processing time and costs and the accuracy ought to have been 95.47 percent. For example, the adaptive learning pipes and enhanced system of governance can be enhanced in the future other than these functions. Such deliverables include novelty of AI to enhance mortgage technology, and developing scalable and intelligent financial systems to manage document-based financial processes.

REFERENCES

[1] Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., Lu, Y., Florencio, D., Zhang, C., Che, W., Zhang, M., & Zhou, L. (2020). LayoutLMV2: Multi-modal pre-training for Visually-Rich Document Understanding. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2012.14740

[2] Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., & Park, S. (2021). OCR-free Document Understanding Transformer. *arXiv (Cornell*

University). https://doi.org/10.48550/arxiv.2111.15664

[3] CVIT, IIIT Hyderabad, India & Computer Vision Center, UAB, Spain. (2020). DocVQA: A Dataset for VQA on Document Images. *DocVQA: A Dataset for VQA on Document Images*. https://openaccess.thecvf.com/content/WACV2021/papers/Mathew_DocVQA_A_Dataset_for_VQA_on_Document_Images_WACV_2021_paper.pdf

[4] Visalli, F., Patrizio, A., Lanza, A., Papaleo, P., Nautiyal, A., Pupo, M., Scilinguo, U., Oro, E., Ruffolo, M., altilia.ai, & High Performance Computing and Networking Institute of the National Research Council (ICAR-CNR). (2023). Building a platform for intelligent document processing: opportunities and challenges. In *Ital-IA 2023: 3rd National Conference on Artificial Intelligence*. CEUR Workshop Proceedings. http://ceur-ws.org

[5] Zheng, H., Wang, S., & Huang, L. (2024). A comprehensive survey on Document-Level information extraction. In Association for Computational Linguistics, *Proceedings of the Workshop on the Future of Event Detection (FuturED)*. https://aclanthology.org/2024.futured-1.6.pdf

[6] Abdallah, A., Eberharter, D., Pfister, Z., & Jatowt, A. (2024). A survey of recent approaches to form understanding in scanned documents. *Artificial Intelligence Review*, *57*(12). https://doi.org/10.1007/s10462-024-11000-0

[7] Subramani, N., Matton, A., Greaves, M., & Lam, A. (2020). A survey of deep learning approaches for OCR and document understanding. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2011.13534

[8] Pramanik, S., Mujumdar, S., & Patel, H. (2020). Towards a Multi-modal, Multi-task Learning based Pre-training Framework for Document Representation Learning. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2009.14457

[9] Amjad, A., Sthapit, S., & Syed, T. Q. (2025, May 16). An agentic system with reinforcement-learned subsystem improvements for parsing form-like documents. arXiv.org. https://arxiv.org/abs/2505.13504

[10] Giovannini, S., Coppini, F., Gemelli, A., & Marinai, S. (2025). BoundingDocs: a Unified Dataset for Document Question Answering with Spatial Annotations. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2501.03403