Anuj Singh[1]

Dr.Pratap Singh[2]

Improving Financial Invoice Workflows with RPA and OCR Using Multimodal Techniques

Journal of Electrical Systems

Abstract

The global financial process is labor-intensive and time-consuming because it relies heavily on written documents and physical work. By combining CV, NLP techniques, and RPA, we have moved towards higher automation to address this issue. Tasks like document categorization and key information extraction fit well with these advanced solutions. However, challenges arise when analyzing text-rich document images, and a large training dataset is needed to process bilingual documents. To automate business processes using practical financial document models, particularly for banking operations, this study introduces an intelligent document processing framework. This framework utilizes a multimodal approach that combines traditional RPA with a pre-trained deep learning model. The proposed system can effectively analyse multilingual documents and is designed to perform categorization and key information extraction with less training data. Extensive studies using images of Indian financial documents were conducted to assess the framework's effectiveness. The results indicate that the multimodal approach is better at interpreting financial documents, and precise labeling can enhance performance by as much as 15%. This framework has greatly improved the automation and optimization of financial document processing.

Keywords: OCR, financial documents, RPA, multimodal, Improving OCR

I. INTRODUCTIONS

The financial sector is increasingly focused on automating business processes to improve productivity and efficiency. This trend is evident in how many banks are adopting technologies like Robotic Process Automation (RPA) and Artificial Intelligence (AI). A specific example of this is the use of intelligent document processing (IDP) systems, which simplify the manual entry of business documents into back-office systems. Recent studies show that finance departments could save over 25,000 hours of rework each year by automating 29% of their operations with RPA. This moves to automation boosts efficiency and allows financial institutions to redirect resources to more strategic tasks, ultimately enhancing their overall performance and competitiveness in the market [8].

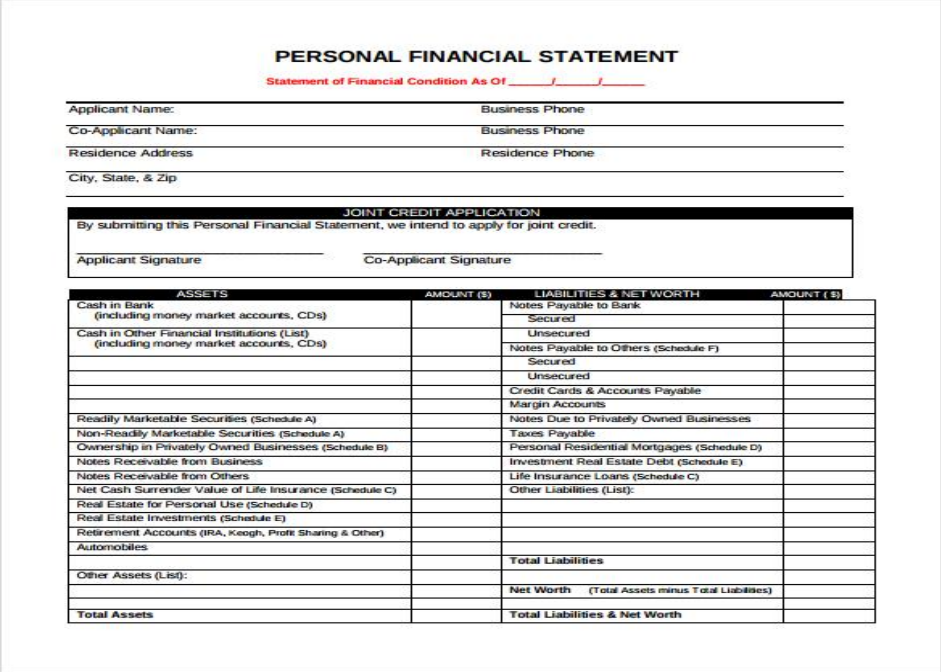
RPA systems perform a variety of repetitive tasks by interacting with other systems to gather and generate data. RPA often uses optical character recognition (OCR) to identify specific areas in structured document images to extract information. By enabling quicker responses, these frameworks significantly speed up business processes and offer greater scalability and flexibility [9]. However, unstructured documents, like salary transaction statements, trade transaction confirmations, and tax slips, still make up a large portion of the banking industry and pose their own challenges [10]. This is illustrated in Fig 1, which shows relevant details of Indian Financial Documents.

Traditional RPA has evolved into hyper-automation through the integration of natural language processing (NLP) and computer vision (CV) techniques [11,12], thanks to advances in AI. Key information is automatically extracted through key information extraction (KIE) and entered into back-office systems. For example, invoice images provide important details such as "total price," "tax amount," and "change amount." Most methods involve using OCR engines to convert documents into text and applying named entity recognition (NER) and other NLP techniques. These systems carry out KIE with minimal labeled data, thanks to pre-trained deep learning models for visual document analysis [13]. However, they require a significant amount of training data for multilingual document analysis and struggle with text-rich document images.

[1*]Scholar, Quantum University Roorkee, Uttarakhand, India. Email- anuj.sre22@gmail.com

[2]Associate Professor, Quantum University Roorkee, Uttarakhand, India. Email- partap.cse@quantumeducation.in

This research presents an IDP framework that connects traditional RPA with pre-trained deep learning models to extract key information from actual Indian financial document photos. This framework allows for the classification of documents and the identification of important information and relationships. It uses a multilingual model that has been trained on natural language data and a wide range of real-world document images. This is the first time a pre-trained multilingual model has been applied in traditional banking processes like data warehousing and RPA for interpreting visually rich documents.



PERSONAL FINANCIAL STATEMENT

Statement of Financial Condition As Of _____/_____/_____

Applicant Name:	Business Phone
Co-Applicant Name:	Business Phone
Residence Address	Residence Phone
City, State, & Zip	

JOINT CREDIT APPLICATION

By submitting this Personal Financial Statement, we intend to apply for joint credit.

Applicant Signature _____ Co-Applicant Signature _____

ASSETS	AMOUNT ($)	LIABILITIES & NET WORTH	AMOUNT ($)
Cash in Bank (including money market accounts, CDs)		Notes Payable to Bank	
		Secured	
Cash in Other Financial Institutions (List) (including money market accounts, CDs)		Unsecured	
		Notes Payable to Others (Schedule F)	
		Secured	
		Unsecured	
		Credit Cards & Accounts Payable	
		Margin Accounts	
Readily Marketable Securities (Schedule A)		Notes Due to Privately Owned Businesses	
Non-Readily Marketable Securities (Schedule A)		Taxes Payable	
Ownership in Privately Owned Businesses (Schedule B)		Personal Residential Mortgages (Schedule D)	
Notes Receivable from Business		Investment Real Estate Debt (Schedule E)	
Notes Receivable from Others		Life Insurance Loans (Schedule C)	
Net Cash Surrender Value of Life Insurance (Schedule C)		Other Liabilities (List):	
Real Estate for Personal Use (Schedule D)			
Real Estate Investments (Schedule E)			
Retirement Accounts (IRA, Keogh, Profit Sharing & Other)			
Automobiles			
		Total Liabilities	
Other Assets (List):			
		Net Worth (Total Assets minus Total Liabilities)	
Total Assets		**Total Liabilities & Net Worth**	

Fig 1. Sample of Unstructured Indian Registration documents

1.1 Our Work

This research introduces a new framework for processing document photos for Key Information Extraction (KIE). It combines traditional Robotic Process Automation (RPA) with pre-trained deep learning models based on real-world Indian financial document images. The method processes document pictures through a database using RPA, encodes them, and tests their effectiveness against language models like XLM-RoBERTa and InfoXLM. An ablation study shows that labeling both "key-value" pairs, instead of just values, can boost accuracy by up to 10%. The study compares the performance of KIE across four models, with a special focus on the effectiveness of "key-value" labeling. It also demonstrates that this approach can cut business processing time for some financial tasks by over 30%, showing the efficiency improvements gained.

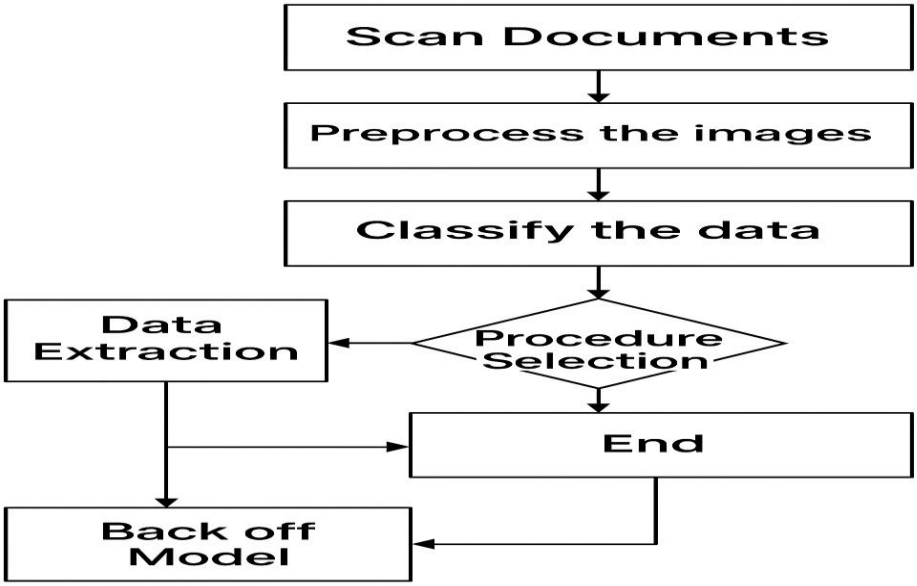


Fig 2. Flowchart of Business Process

In Fig 2, we see a typical business process flowchart for handling Indian financial images or scanned documents. This is done to improve the understanding of the entire application. It uses robot processing automation and optical character recognition.

The organizing of our research is explained in Section 2, which serves as a literature review. Section 3 provides the complete procedure for implementing the framework and methods. Section 4 discusses the results and their experimental implications. Finally, Section 5 summarizes the conclusions of the research.

II. LITERATURE REVIEW

This section discusses important research on using AI in document image processing, especially on pre-trained models for understanding multilingual document images. We focus on two main topics: trained methods for multilingual document interpretation and intelligent document processing systems.

2.1 Business Process Automation

AI developments have made it easier to digitize data across various industries, including manufacturing, finance, insurance, and healthcare [14,17-19]. Many attempts have been made to automate these processes using deep learning or machine learning techniques. For example, Baidya [17] suggested an automated business process for classifying unstructured documents that employs RPA and a support vector machine classifier. Roopesh et al. [18] created an intelligent system that uses AI and RPA to monitor emails and attachments, categorize resumes with deep learning models, and extract relevant data from the resumes. Their method relies on the LSTM-CRF model and uses OCR-extracted text as input for named entity recognition (NER) and bidirectional long short-term memory (LSTM). Their approach is based on using deep learning techniques to improve the accuracy and efficiency of insurance document categorization. In their study, they introduced a new multimodal binary classification method designed specifically for this task, incorporating transfer learning as a key part.

By using the pre-trained Bidirectional Encoder Representations from Transformers (BERT) model along with VGG-16, they successfully extracted both text and visual information from document images, enhancing the input data for the classification process. By combining textual and visual features, they presented a deep learning method that integrates Layout LM with residual networks [21]. This technique achieves the best performance in identifying breaking points in visually rich document sequences.

2.2 Multilingual Training Methods

The methods involve training transformers on a range of language data, which has been crucial for commercial digital changes [21]. BERT and its multilingual version, mBERT, were introduced after being pre-trained on

104 languages using Wikipedia data and a masked language modelling method [22-23]. To create cross-language models (XLM), [23-24] presented two machine learning approaches: one that relies on monolingual data and another that uses parallel collections. They developed a multilingual modelling technique that maintains performance for each language. They also introduced a sequential-to-sequential denying auto-encoder, known as the bidirectional auto-regressive transformer, which was already trained on large amounts of monolingual data in various languages [25].

2.3 Deep Learning and RPA Pre-Trained IDP Framework:

We offer a new intelligent document processing (IDP) model that combines traditional robotic process automation (RPA) with pre-trained deep learning models. This framework connects standard RPA with modern AI methods. It is specifically designed to extract important information from real Indian financial documents.

2.4 Multilingual Model for Financial Documents:

Our fundamental research uses a multilingual model. Earlier research mainly focused on single-language or cross-language models for industries like banking, healthcare, insurance, and manufacturing. This approach gives global financial organizations a more complete and flexible solution, as it can handle financial documents written in various languages.

2.5 Key-Value Pairs:

To explore how identifying both "key-value" pairings impacts document understanding, we conduct an ablation study. Compared to labeling only values, this method shows an accuracy improvement of up to 10%. This highlights the importance of thorough labeling for enhancing the efficiency of document processing systems.

2.6 Reduction in Business Processing Time:

Our framework shows a significant reduction in business processing time. It surpasses thirty percent for several financial processes. This boost in efficiency highlights the benefits of our method. It serves as an important resource for financial organizations that want to reduce manual work and simplify procedures.

III. OUR PROPOSED WORK

Our proposed method for evaluating financial document images is detailed in this section and illustrated in Fig3. The main goals of the framework are layout analysis, extracting key information, and document categorization. Transformer-based models are effective due to their self-attention mechanism. This feature allows flexible input component weighting, which helps highlight important elements for each task. These models are fine-tuned for specific tasks and have been pre-trained on large datasets, making them good at handling variable-length inputs and adapting to new data. They also perform well even with a small amount of task-specific training data. Additionally, transformer-based models can manage multiple tasks simultaneously. This capability helps them gain more general representations and transfer knowledge between tasks efficiently.

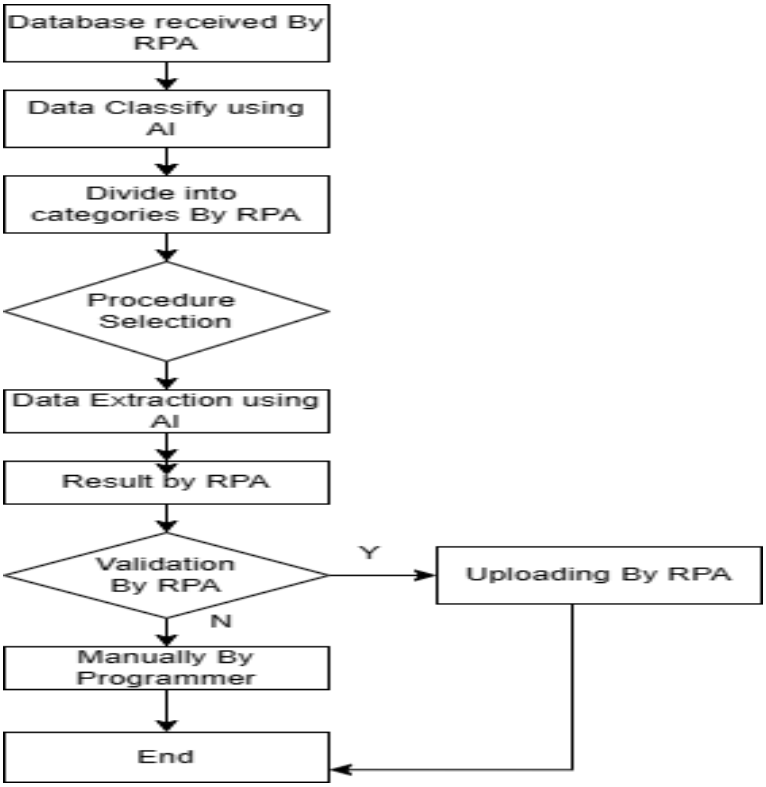


Fig 3. Our Scan Proposed Indian Financial Document Processing System

3.1 Entire Package

Our architecture has three main parts: RPA, AI, and human intervention. It integrates AI with an automated robotic process automation (RPA) bot. To start the process, a user scans printed documents into a database. RPA then retrieves the documents as image files. A multimodal strategy replaces human document categorization with a well-tuned LayoutXLM model, which improves accuracy and reduces costs.
After classifying each page of the document, RPA processes it and sends the images to KIE-tuned models based on their class. These models extract important information by fine-tuning them according to set criteria for each type of document. For example, from a business registration certificate, the model extracts key information like the company's name, the representative's name, the starting date, and the registration number.

Once the data is extracted, it is converted into an Excel file with set classifications. RPA then notifies the user by email when the extraction process is complete. Finally, the user performs additional manual checks and corrections to ensure the retrieved data is accurate.

3.2 Handling Informal Financial Documents

The information describes how to use Layout XLM, a transformer-based model, for unstructured multilingual document images in document categorization and Key Information Extraction (KIE). Layout XLM has undergone specialized training to understand documents by focusing on both layout details and image features. This method is different from traditional pre-trained language models like mBERT and Info LM, which rely mainly on diverse multilingual text datasets for training.
To effectively capture the complex patterns in document layouts, Layout XLM uses a transformer-based architecture. This architecture includes a stack of multi-head self-attention layers and a feed-forward network, featuring a self-attention mechanism that is aware of spatial relationships. The training set for Layout XLM consists of eleven million multilingual document images from the IIT-CDIP dataset, which are detailed in Fig 4.
Layout XLM is a valuable tool for processing various document formats across many languages and fields. It excels in tasks such as document classification and KIE by using both textual and visual information.

3.2.1 LayoutXLM pretrains using four different kinds of embeddings

The information explains the methods for embedding and processing document images that the LayoutXLM model uses. Text embedding involves adding special tokens, such as [PAD], [SEP], and [CLS], into OCR text sequences to help with tokenization and maintain consistent sequence length. By using the document image as input and updating the ResNeXt-FPN output feature map through backpropagation, visual embedding lets the model extract visual features. Additionally, positional embedding includes both 1D and 2D techniques. The 2D positional embedding captures the positions of text sequences within the document, which shows the spatial layout. In contrast, 1D positional embedding reflects the reading order of text sequences.
When combined, these embedding methods boost LayoutXLM's ability to manage the visual and linguistic aspects of document images. LayoutXLM improves the understanding and analysis of document layouts by merging text and visual embeddings. This results in better performance in tasks like Key Information Extraction (KIE) and Document Image Classification (DIC). Overall, LayoutXLM is a powerful tool for document comprehension tasks across various languages and fields due to the integration of different embedding techniques, allowing it to interpret multilingual document images and reliably extract essential information.

3.2.2 Three pretraining procedures are used in the training of LayoutXLM

Masked Visual-Language Modeling involves randomly masking text tokens and then recovering them. Text-picture Alignment helps understand the spatial relationship between text and image coordinates. Text-picture Matching focuses on the connection between text and image information. We adjust all features in financial document images within LayoutXLM to improve its effectiveness for document categorization and Key Information Extraction (KIE). In the KIE task, LayoutXLM employs sequential labeling to identify and classify each entity type by predicting {B, O} tags for each token. The model uses the [CLS] token and its representation to predict class labels for documents. This approach enables precise classification and extraction of vital information from multilingual, unstructured financial records.

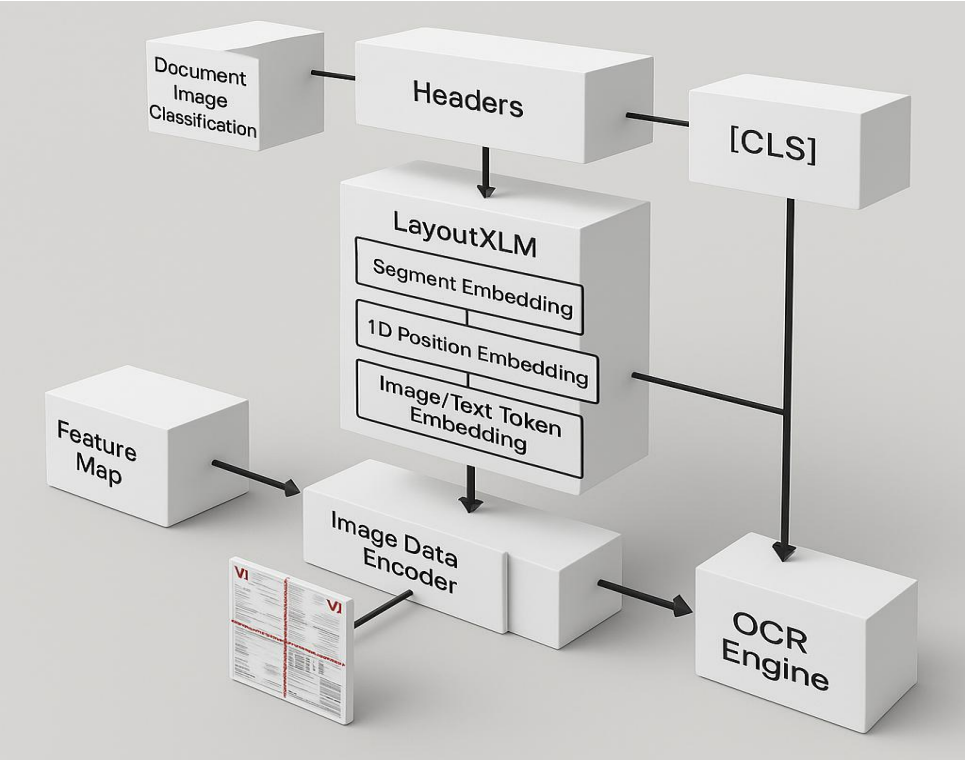


Fig 4. Document Image Classification Architecture Model

IV. RESULT AND EXPERIMENT

The experiments conducted to achieve two main goals are described in this section: (1) Key Information Extraction (KIE) and (2) Document Image Classification (DIC). We used a 14-layer transformer encoder with 14 attention heads and a hidden size of 779 to fine-tune the base models. ResNeXt-101-FPN was used to build Layout XLM's optical backbone network. Since the larger models have not been made public, we assessed the proposed framework by comparing it with several pre-trained language models, including the base size models of mBERT and Layout XLM. Each experiment ran in the Python 11.3 environment on an Asus Core i5 10th Gen with a 1TB SSD, 8GB RAM, and a single NVIDIA GPU. The Hugging Face model hub provides the pre-trained models for download. They were trained using CUDA 12.6 and PyTorch 1.9.1. We used standard evaluation measures for Named Entity Recognition (NER) and document classification, including macro-averaged precision, macro-averaged recall, macro-averaged F1 score, and accuracy, to assess how well the different models performed on the two tasks. Since most entities in the KIE task fall under the "other" class, the macro-averaged scores are not affected by this fact. The macro-averaged scores are calculated as the arithmetic mean of each per-class score. The following equations describe these parameters:

$$Accuracy = \frac{TP + TN}{T\,P + T\,N + F\,P + FN} \quad (I)$$

$$Precision = \frac{TP}{TP + FP} \quad (II)$$

$$Recall = \frac{TP}{TP + FN} \quad (III)$$

$$F1 - score \; = 2 \times \frac{(Precision \; \times Recall)}{(Precision \times Recall)} \quad (IV)$$

The results of the studies show that our multimodal strategy, which uses Layout XLM to combine text and layout understanding, provides a significant advantage in KIE and DIC tasks, particularly for complex document images.

4.1. Dataset

Images of business documents obtained between March 1, 2022, and August 31, 2022, using our current products were used in this study. The documents were scanned as images, and we used the NAVER CLOVA OCR engine to extract the text. We tagged entities for extraction in each document for the KIE job, following the format of the CORD and FUNSD datasets. One person conducted the inspection, and two others completed the labeling process. Most financial papers require different types of document images, such as semi-structured or unstructured. For manual data extraction from banks, the dataset needs to include key-value pairs.

As a result, we used two datasets for KIE: (1) Bill of Lading (BL) and (2) Business Registration Certificate (BRC). These financial datasets meet the earlier mentioned requirements and include enough key-value pairs.

We trained and assessed our method using pictures of Indian company registration certificates as a fine-tuning dataset for the KIE task. This collection includes 18,326 fully annotated items and 252 scanned document images. We divided the BRC dataset into five categories: 1,850 others, 2,541 document information, 3,189 tax information, 10,176 company information, and 570 alcohol sales information. Most of these subclasses are organized as linked semantic entities. We refer to the related semantic entity pairs as "key" and "value," marking them with an asterisk (*).

We also trained and assessed our proposal for the KIE task on scanned financial document images with multilingual text (English, Indian, Chinese, and others). This collection contains 60,610 fully annotated items and 230 BL document images. Data from the BL dataset are shown in Table 1. The five categories include 888 document information, 2,922 shipper information, 4,989 consignee information, 3,092 locations, and 48,719 others. From these five categories, we created 21 subclasses. An asterisk (*) marks the related semantic items, similar to what we did in the BRC dataset.

Table 1: BL Dataset Statistics

Superclass	Number of Items	Subclass
Document Information	888	4
Consignee Information	4,989	3
Locations	3,092	8
Shipper Information	2,922	4
Others	48,719	2

4.2. Extracting Essential Facts

The Key Information Extraction (KIE) job results and analysis using several transformer-based models are shown in this section. These models perform well on various NLP tasks with minimal task-specific fine-tuning. They are specifically designed for NLP tasks and have been pre-trained on large multilingual datasets. An explanation of the already trained models is provided below.

a) **XLM-RoBERTa:** Already trained model across more than 100 languages on a 3 TB text amount.
b) **mBERT:** Developed using a 105GB corpus of textual dataset in 110 distinct languages.
c) **Layout XLM:** Using a dataset of 40 million document visuals in 61 different languages, this system was able to extract layout information as well as text from document images.
d) **Info XLM:** It has been already trained on a large corpus in more than 104 languages and is intended to handle jobs involving both text-based as well as structured data-based natural language processing.

We examined these four models to assess KIE's performance on the BRC and BL datasets. We used the open pre-trained models mentioned in relevant research for each experiment. To evaluate the effects of labeling "key" and "value" entities compared to just "value" entities, we also conducted an ablation study. This was motivated by the fact that benchmarks and related research generally follow one of these two approaches.
To see how layout features affected the model's performance, we changed the BRC and BL datasets' "key" entities to "other" entities. This helped us understand how much layout information impacts overall efficiency, considering Layout XLM's ability to interpret layout as well as text. In terms of overall accuracy and macro-averaged scores, Layout XLM consistently outperformed other models across all datasets, as shown by the results in Table 2. Specifically, Layout XLM achieved a macro-F1 score of 0.8874 on the BL dataset, outperforming the other models. This demonstrates that our proposed method, which uses Layout XLM, works particularly well for document images that contain a lot of text.

Table 2: Performance Metrics for KIE Task

Model	Dataset	Macro-F1 Score	Accuracy
mBERT	BRC	0.8123	0.8437
XLM-RoBERTa	BRC	0.8245	0.8561
InfoXLM	BRC	0.8357	0.8634

Model	Dataset	Macro-F1 Score	Accuracy
LayoutXLM	BRC	0.8712	0.8920
mBERT	BL	0.7998	0.8215
XLM-RoBERTa	BL	0.8145	0.8394
InfoXLM	BL	0.8289	0.8472
LayoutXLM	BL	0.8874	0.9012

Sampled results from the BRC and BL datasets are shown in Fig 5. Three versions of each image are included. The first version displays full results with different colored text boxes that indicate headers, keys, values, and other information. The second version shows models trained without key labels. These findings reveal that some "value" labels are often misclassified as "other" labels by models that do not use "key" labels. This highlights the importance of using key-value pairs and layout information to improve KIE accuracy in document images. LayoutXLM's strong performance underscores the benefits of combining text and layout knowledge in analysing financial document images.

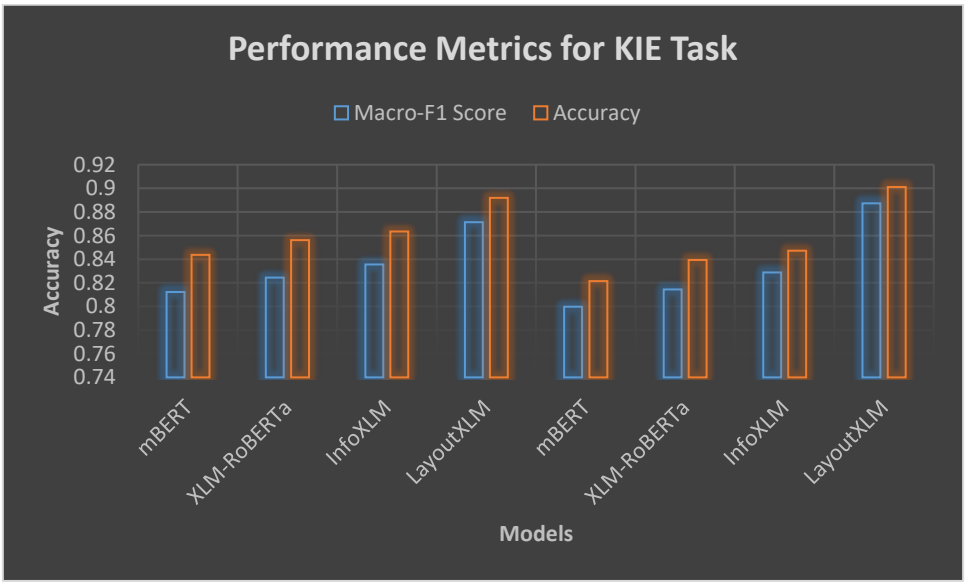


Fig 5. Shown relation of BRC and BL datasets model between accuracy

Furthermore, as Table 3 shows, Layout XLM performed the best across all datasets in the ablation study. According to the BL dataset, the average F1-score dropped by 11% when evaluating text-only pre-trained models. All models experienced a significant drop in recognition performance for the "value" labels. This suggests that extracting these labels relies on key-value pair annotations. However, Layout XLM showed remarkable resilience to these changes. Compared to other models, Layout XLM maintained a similar average F1 score even after removing the "key" labels. More specifically, recognition results for only the "value" labels declined by 1.5%. This small drop indicates that removing key labels has little effect on Layout XLM's ability to understand and extract information from documents containing key-value pairs.

Table 3: Macro-Averaged F1 Scores

Model	Dataset	With Keys	Without Keys	Decrease (%)
mBERT	BL	0.8764	0.7748	11.6%
XLM-RoBERTa	BL	0.8852	0.7811	11.7%
InfoXLM	BL	0.8783	0.7759	11.7%
LayoutXLM	**BL**	**0.8874**	**0.8741**	**1.5%**

According to these results, labeling "key-value" pairs is particularly helpful for understanding text in structured data sets, such as BL and BRC. Additionally, our proposed method performed better than document images with only "value" labels. The findings show that Layout XLM works well in both text-based and layout-based document comprehension tasks due to its dual pre-training on text and layout information. This extensive pre-training allows Layout XLM to manage a variety of document structures while maintaining strong performance across many situations.

4.3. Classification of Document Images

This section compares three transformer-based models that have been trained to a traditional Convolutional Neural Network (CNN) model to evaluate performance in document depiction classification. To conduct a thorough analysis, we have selected two popular CNN models for the comparison experiments: VGG-16 and InceptionResNet-V2. We also include a summary of the pre-trained models.
VGG-16: A widely recognized CNN model known for its ease of use and efficiency in image classification.
InceptionResNet-V2: A convolutional neural network model uses residual connections along with the Inception architecture to capture complex patterns and relationships in images. We tested the same datasets, SWT and SD, that were used for the Key Information Extraction (KIE) task to evaluate how well the models classify document images.
We compared the performance of Layout XLM, m-BERT, XLM-Ro-BERTa, and the selected CNN models. The results of our proposed method and the models' performances are shown in Table 4 and Fig 6. Layout XLM ranked second on the SD dataset but had the best classification results on the SWT dataset. Transformer-based models are usually much larger than traditional CNN models like VGG-16 and InceptionResNet-V2. Their larger size allows them to capture more complex connections and patterns in the data.
CNN models that relied only on visual features performed significantly worse on the SWT dataset because every document image includes a table and looks similar. However, CNN models did better on the SD dataset, which includes images with diverse visual shapes.

Table 4: Document Image Classification Performance

Model	SWT Dataset	SD Dataset
VGG-16	0.7823	0.8156
InceptionResNet-V2	0.8052	0.8327
mBERT	0.8159	0.8253
XLM-RoBERTa	0.8126	0.8374
LayoutXLM	0.8294	0.8291

4.4 Comparative Results of Experiment:

These findings show that while transformer-based models like Layout XLM, m-BERT, and XLM-RoBERTa often outperform CNN models in some situations, CNN models still perform well, especially in datasets with different visual forms. This highlights the importance of choosing the right model based on the needs of the task and the features of the dataset, as shown in Fig 6.

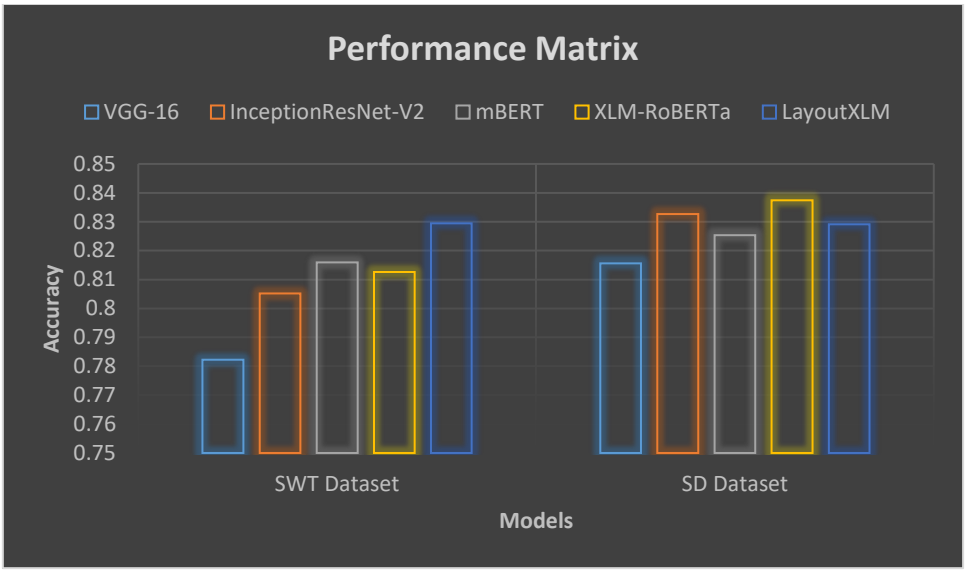


Fig 6. Performance matrix evaluation with SWT dataset and SD dataset

V. CONCLUSION AND FUTURE WORK

This study introduced a new Intelligent Document Processing (IDP) approach using a multimodal method to automate Key Information Extraction (KIE) from real Indian financial document images. It combines standard Robotic Process Automation (RPA) with pre-trained deep learning models. We used RPA for document encoding and retrieval, then applied the optimized Layout XLM model for KIE and document categorization tasks. Our proposed VrDU pre-trained model performed better than leading language models like XLM-RoBERTa and Info XLM in comparison tests. This shows its superior accuracy in KIE and document categorization tasks, which are crucial for banking operations. Our ablation study highlighted the importance of labeling both "key-value" pairs, showing a 10% increase in accuracy compared to labeling only values. While our technique demonstrates strong document understanding, challenges remain, especially with token classification performance due to the limited availability of Indian document images in the pre-trained data. To address this issue, future work will focus on gathering and pre-training more images of financial documents that specifically improve model performance on Indian papers. Additionally, we aim to explore how to extend our system to support languages beyond Indian and how to tackle challenges to ensure effective multilingual document processing.

Our future research plan includes several key areas to pursue. First, to enhance the model's understanding of Indian documents, we will expand our dataset by collecting more diverse financial document images and conducting thorough pre-training. We also intend to explore transfer learning techniques to adapt our system for languages other than Indian, ensuring its usefulness across different language contexts. Furthermore, we plan to refine the model architecture and add domain-specific features to boost token classification performance. Finally, we seek to partner with industry players to implement and evaluate our framework in real banking environments, gathering feedback to improve its effectiveness and efficiency.

REFERENCE

[1] Maqsood, Haider, MuazzamMaqsood, Sadaf Yasmin, Irfan Mehmood, Jihoon Moon, and Seungmin Rho. "Analyzing the stock exchange markets of EU nations: A case study of brexit social media sentiment." Systems 10, no. 2 (2022): 24.

[2] Jabeen, Ayesha, Muhammad Yasir, Yasmeen Ansari, Sadaf Yasmin, Jihoon Moon, and Seungmin Rho. "An Empirical Study of Macroeconomic Factors and Stock Returns in the Context of Economic Uncertainty News Sentiment Using Machine Learning." Complexity 2022 (2022).

[3] Zaib, Aurang, Muhammad Yasir, and Mohammad Usman. "The Role of Urdu Literature in the Independence Struggle of the Subcontinent." Pakistan Languages and Humanities Review 7, no. 1 (2023): 421-433.

[4] Al Ansari, Yasmin, HalimeShahwan, and Bruno Ramos Chrcanovic. "Diabetes mellitus and dental implants: a systematic review and meta-analysis." Materials 15, no. 9 (2022): 3227.

[5] Schlegel, Dennis, and Jonathan Wallner. "Research on robotic process automation: structuring the scholarly field." In Business advancement through technology volume II: The changing landscape of industry and employment, pp. 19-45. Cham: Springer International Publishing, 2022.

[6] Tang, Qing, YoungSeok Lee, and Hail Jung. "The Industrial Application of Artificial Intelligence-Based Optical Character Recognition in Modern Manufacturing Innovations." Sustainability 16, no. 5 (2024): 2161.

[7] Jain, Arushi, ShubhamPaliwal, Monika Sharma, LovekeshVig, and Gautam Shroff. "SmartFlow: Robotic Process Automation using LLMs." arXiv preprint arXiv:2405.12842 (2024).

[8] Shailaja, P., and V. Naga Lakshmi Ponnada. "A REVIEW ON ROBOTIC PROCESS AUTOMATION IN PHARMACEUTICAL INDUSTRY." (2024).

[9] Namaz, Isabela. "Refactoring test automation framework using optical character recognition." (2024).

[10] Auer, Thomas, and Christian Schieder. "No Need to Cry over Spilt Milk: A Workflow for Regenerating Graph Data Using Robotic Process Automation." In International Conference on Design Science Research in Information Systems and Technology, pp. 247-261. Cham: Springer Nature Switzerland, 2024.

[11] Li, Bohan, Yonghua Shi, and Zishun Wang. "Penetration identification of magnetic controlled Keyhole Tungsten inert gas horizontal welding based on OCR-SVM." Welding in the World (2024): 1-12.

[12] Datta, Arkajit, TusharVerma, and Rajat Chawla. "AUTONODE: A Neuro-Graphic Self-Learnable Engine for Cognitive GUI Automation." arXiv preprint arXiv:2403.10171 (2024).

[13] Deepa, S., VivekDuraivelu, BalamuruganEaswaran, R. Suguna, and H. Aparna. "Enhancing Road Safety with Real-Time Helmet Detection and E-Challan Issuance using YOLO and OCR." In 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS), pp. 644-649. IEEE, 2024.

[14] Cho, Seongkuk, Jihoon Moon, Junhyeok Bae, Jiwon Kang, and Sangwook Lee"A Framework for Understanding Unstructured Financial Documents Using RPA and Multimodal Approach" Electronics 12, no. 4: 939., 2023.

[15] Jacobus, Michael, and Martin Schneider. "Robotic Process Automation in Financial Institutions." Digital Project Practice for Banking and FinTech (2024).

[16] Vijayalakshmi, K., M. Dhanamalar, Vijayalakshmi A. Lepakshi, and SonamJamtsho. "Smart Checkpoint Management System for Automatic Number Plate Recognition in Bhutan Vehicles Using OCR Technique." SN Computer Science 5, no. 5 (2024): 579.

[17] Farinha, Diogo, Ruben Pereira, and Rafael Almeida. "A framework to support Robotic process automation." Journal of information technology 39, no. 1 (2024): 149-166.

[18] Haavisto, William. "Automating the Certificate Verification Process." (2024).

[19] Su, Guanqun, Shuai Zhao, Tao Li, Shengyong Liu, Yaqi Li, Guanglong Zhao, and Zhongtao Li. "Image format pipeline and instrument diagram recognition method based on deep learning." Biomimetic Intelligence and Robotics 4, no. 1 (2024): 100142.

[20] Tiron-Tudor, Adriana, Ramona Lacurezeanu, Vasile Paul Bresfelean, and AdelinaNicoletaDontu. "Perspectives on How Robotic Process Automation Is Transforming Accounting and Auditing Services." Accounting Perspectives 23, no. 1 (2024): 7-38.

[21] Manzoor, Shahid, Nimra Wahab, and MKA Ahamad Khan. "An Improved Algorithm for Optical Character Recognition using Graphical User Interface Design."

[22] Beerbaum, Dirk. "Artificial Intelligence (GAI) Ethics Taxonomy-Generative GAI applying Chat GPT for Robotic Process Automation (GAI-RPA) as business case."

[23] Varma, Sandeep, ShivamShivam, Soumya Deep Roy, and Biswarup Ray. "A Rule-Based Expert System for Automated Document Editing." In International Conference on Computing and Information Technology, pp. 85-94. Cham: Springer Nature Switzerland, 2024.

[24] Seiler, Andreas. The Impact of Order Processing Automation on Customer Care and Sales Logistics. SAGE Publications: SAGE Business Cases Originals, 2024.

[25] Norbert, G. A. L., VasileStoicu-Tivadar, and G. A. L. Emanuela. "Robotic Process Automation Based Data Extraction from Handwritten Medical Forms." In Telehealth Ecosystems in Practice: Proceedings of the EFMI Special Topic Conference 2023, vol. 309, p. 68. IOS Press, 2023.

[26] Shidaganti, Ganeshayya, R. Sanjana, K. Shubeeksh, VR Monish Raman, and ValmeekiThakshith. "Chatgpt: Information retrieval from image using robotic process automation and ocr." In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1264-1270. IEEE, 2023.

[27] William, P., Siddhartha Choubey, AbhaChoubey, and Gurpreet Singh Chhabra. "Evolutionary Survey on Robotic Process Automation and Artificial Intelligence: Industry 4.0." Robotic Process Automation (2023): 315-327.

[28] Malathi, T., DiwaanChandar, S. Nithish, V. Niranjan, and A. K. Swashthika. "Analyzing and Experimenting Open Source OCR Engines in RPA with Levenshtein Distance Algorithm."

[29] Tang, Qing, YoungSeok Lee, and Hail Jung. "The Industrial Application of Artificial Intelligence-Based Optical Character Recognition in Modern Manufacturing Innovations." Sustainability 16, no. 5 (2024): 2161.

[30] Gupta, Abhishek, Pranav Soneji, and NikhitaMangaonkar. "Robotic process automation powered admission management system." In Inventive Computation and Information Technologies: Proceedings of ICICIT 2022, pp. 169-178. Singapore: Springer Nature Singapore, 2023.

[31] Ribeiro, Jorge, Rui Lima, Tiago Eckhardt, and Sara Paiva. "Robotic process automation and artificial intelligence in industry 4.0–a literature review." Procedia Computer Science 181 (2021): 51-58.

[32] Sharma, Vijay Kumar, Swati Sharma, MukeshRawat, and Ravi Prakash. "Adaptive Particle Swarm Optimization for Energy Minimization in Cloud: A Success History Based Approach." In Towards the Integration of IoT, Cloud and Big Data: Services, Applications and Standards, pp. 115-130. Singapore: Springer Nature Singapore, 2023.

[33] Sharma, Vijay Kumar, Md Iqbal, and Krishna Mohan Pandey. "A Unique Metaheuristic Algorithm for Human Urbanisation." In 2023 International Conference on Device Intelligence, Computing and Communication Technologies,(DICCT), pp. 468-471. IEEE, 2023.

[34] Sharma, Swati, Vijay Kumar Sharma, Rohit Aggarwal, and Rashmi Gupta. "Covid-19 Pandemic Predictive System Using Machine Learning." In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 534-539. IEEE, 2022.

[35] Shukla, Shubham, Ajay Kumar Singh, and Vijay Kumar Sharma. "Survey on Importance of Load Balancing for Cloud Computing." In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 1479-1484. IEEE, 2021.

[36] Sharma, Vijay Kr, Agam Gupta, Abhishek Kumar, and Mayank Kumar Singh. "Scrap Collection System E-Commerce Platform: Online Kabadiwala." In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), vol. 5, pp. 787-790. IEEE, 2024.

[37] Gupta, Muskan, Prakarti Singh, and Vijay Kumar Sharma. "Loan eligibility prediction model using machine learning algorithms." In Artificial Intelligence, Blockchain, Computing and Security Volume 2, pp. 20-26. CRC Press, 2024.

[38] Sharma, Vijay Kumar, Vimal Kumar, Shashwat Pathak, Naman Malik, and Rachit Arora. "Image Web Crawler Towards Machine Learning." In Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020, pp. 913-921. Springer Singapore, 2021.

[39] Sharma, Swati, Vijay Kumar Sharma, Vimal Kumar, and Umang Arora. "Machine Learning Application: Sarcasm Detection Model." Artificial Intelligence for a Sustainable Industry 4.0 (2021): 125-138.

[40] Arora, Shreya, Tushar Bhatia, Stuti Rastogi, and Vijay Kumar Sharma. "Comparison of Various Classifiers for Movie Data." In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 138-140. IEEE, 2021.