Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Probabilistic Estimation and Error Bounds in AI-Based OCR Systems for Enterprise Finance

Ranadheer Reddy Charabuddi [1]
[1] *Avventis Inc, USA. Email:ranadheer30@gmail.com*

ARTICLE INFO

Received: 18 June 2025

Revised: 27 Jul 2025

Accepted: 07 Aug 2025

ABSTRACT

Optical Character Recognition (OCR) has become essential for automating document workflows in enterprise finance, handling tasks such as invoice processing, tax extraction, and compliance reporting. While recent improvements have been made on AI-based OCR systems, few of the current models output calibrated confidence estimates or measure uncertainty, and as a result, they pose critical risks in high-stakes financial applications. These systems normally provide deterministic outputs with no facilities for measuring prediction accuracy, which denies transparency and trust in computerized financial decision-making. This study introduces a new probabilistic OCR paradigm that integrates uncertainty estimation and error-bound modeling into the AI pipeline. Built upon the LayoutLM transformer architecture, the framework employs Monte Carlo Dropout during inference to generate multiple predictions per input, enabling the computation of predictive entropy and confidence intervals for both categorical and numerical fields. The methodology includes preprocessing scanned financial documents from the SROIE v2 dataset, text region segmentation, supervised label alignment, and key-value pairing for structured extraction. The implementation uses PyTorch and HuggingFace Transformers, supported by statistical post-processing to flag uncertain outputs and reduce operational risk. Evaluation results demonstrate high reliability, with the proposed system achieving 99.13% accuracy, a mean confidence interval width of ±1.22 for financial fields, and an expected calibration error of just 2.9%. Approximately 12.4% of predictions are flagged for manual review, effectively balancing automation with oversight. By combining layout-aware modeling with principled uncertainty quantification, the system enhances reliability, explainability, and risk-awareness in enterprise finance, making it a strong candidate for trustworthy financial document automation.

Keywords: AI-Based OCR, Probabilistic Estimation, Error Bounds, Enterprise Finance, Uncertainty Quantification.

INTRODUCTION

In the changing world of digital transformation, businesses are looking more and more towards automated document processing to boost business efficiency, especially in finance-related operations [1]. Optical Character Recognition (OCR) has proven itself to be a core technology for financial document digitization and extraction of structured data from financial documents like invoices, receipts, tax reports, and audit reports [2]. OCR is an electric device which facilitates converting the hand written or printed text and images into a machine-readable digitized format that can be further processed according to the requirement such as indexing, searching, analysis and editing. A robot, as opposed to the human brain, can recognize text characters in data by means of visual images only with the assistance of meaningful algorithms [3]. Therefore, a lot of effort has been directed towards

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

developing procedures that translate the images of documents into machine-readable formats. OCR systems are typically categorized into printed based and hand-based OCR systems. Hand writing text recording is more difficult to identify due to variations in hand writing styles so results are less accurate compared to a printed version. Additionally, OCR is a disaster that can be categorized into online and offline systems. OCR online computers occur in real-time and typically are performed on handwriting in response to offline OCR where the process is carried out on a pre-captured static image [4]. The state-of-the-art OCR techniques accomplished the tuning on printed text extraction utilized on single-image-based documents, but some specialized techniques are introduced to work on videos [5]. The OCR technology has been so flexible that it has opened up many applications including automated data entry, automatic vehicle license plate recognition, automatic reading of postal addresses, bank cheque reading and recognition, intelligent driving assistance and invoice processing. The application examples illustrate the growing importance of OCR to automate and streamline the data extraction processes in a variety of domains [6]. Despite these advances, OCR still suffers particularly with low-quality images, document layout complexity, and multilingual scripts that introduce ongoing research in terms of the effective accuracy and reliability of AI-based models in many real-world scenarios [7]. Traditional OCR systems were rule-based, mainly depending on heuristics, regular expressions and simple pattern matching and did not work very well with a broad range of formats, languages and layouts of documents encountered in enterprise documents. The systems are not very great at distracting backgrounds and smeared scans and fonts, not so great in real-world finance documentations.

With the onset of Artificial Intelligence (AI) and deep learning, the OCR technologies have improved enormously [8]. In corporate finance, where financial compliance and integrity are of the utmost importance, AI-based OCR systems are extremely vital in streamlining tasks such as invoice processing, expense reporting, accounts payable/receivable, and regulatory filings. But financial documents are most likely to carry domain-specific jargon, numeric fields embedded in other text, and tabular structures that pose even to advanced models [9], [10]. In spite of these technological improvements, the essential question continues to be: the reliability and accountability of OCR outputs in enterprise financial operations. Mistakes in text extraction, particularly in money amounts, names of entities, or due dates, can cascade into downstream applications and lead to vast financial and regulatory risks [11]. Furthermore, enterprise finance processes require traceability and confidence factors that are not typically available in available OCR applications. With the onset of AI and deep learning, the OCR technologies have improved enormously. In corporate finance, where financial compliance and integrity are of the utmost importance, AI-based OCR systems are extremely vital. For most scenarios, organizations lack insight into the level of confidence the system has in a given extraction or even what the anticipated margin of error will be [12]. This lack of transparency inhibits trust and constrains adoption in high-stakes settings. Since financial information is often utilized for automated decision-making, audit trails, and regulatory reporting, the lack of quantifiable error measures and confidence estimates becomes a main constraint. Closing this gap is crucial to be able to make sure OCR systems used in enterprise finance can cater to not only automation, but also accountability, explainability, and informed human oversight.

Developments in AI-based optical character recognition (OCR) systems over the past few years have seen the emergence of extremely sophisticated architecture with a remarkable boost in extracting text from complicated documents. Architectures like Convolutional Recurrent Neural Networks (CRNNs) have found extensive use in sequence labelling tasks [13], and Vision Transformers (ViTs) provide strong spatial-textual representation capabilities [14]. Pre-trained multimodal models such as LayoutLM, Donut, and TrOCR have further transformed the scene by combining visual and text embeddings and thus allowing these models to process complex layouts, tables, and forms that exist in actual documents [15], [16]. These advances have delivered remarkable performance on benchmark datasets like FUNSD, SROIE, and DocVQA. Moreover, the advent of self-supervised learning and multimodal learning paradigms has made these systems generalize more across tasks without depending so much on costly annotated data [17], [18]. Yet, even though these technological advances have been made, many of these cutting-edge OCR solutions rely on deterministic processes, providing static outputs without qualifying uncertainty or confidence levels against predictions. Such becomes especially troublesome in enterprise finance use cases,

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

where extracted information certainty matters for subsequent procedures such as automated approvals, financial risk analysis, and regulatory compliance procedures [19]. The lack of calibrated uncertainty estimates poses substantial operational hazards, as the incorrect outputs can quietly spread along decision-making pipelines without triggering any protections or fall-back mechanisms. Also, these models tend to be poor at handling domain-specific terminology, irregular document structure, biased orientations, or out-of-distribution inputs and expose their deficits in dealing with real-world variability. Though some instantiations use post-processing rules or have confidence thresholds for rejecting uncertain responses, such approaches do not have principled probabilistic origins and do not offer solid bounds on prediction error. This lack creates no sound means for companies to evaluate the integrity of AI-derived document interpretation or to measure how uncertainties in OCR outputs can affect large-scale financial calculations and strategic business decisions. In general, while notable advancements have been made in precision and generalizability, existing OCR systems are still not well-positioned to meet the sophisticated challenges of enterprise-wide document understanding from a risk-conscious and explainable perspective. To bridge these essential limitations of existing OCR systems, this work introduces a probabilistically founded framework with the aim to reinforce reliability, interpretability, as well as risk-awareness in enterprise finance. This study plans to create a probabilistically tuned AI-driven OCR framework specifically designed for enterprise financial applications

This project plans to create a probabilistically tuned AI-driven OCR framework specifically designed for enterprise financial applications. The main goal is to incorporate uncertainty estimation and error bound modeling within the OCR process through state-of-the-art probabilistic approaches such as Bayesian deep learning and conformal prediction. The performance will be tested over real-world financial documents to achieve robustness, explainability, and decision-level confidence. With the addition of statistically interpretable error bounds, the new framework will allow enterprise users to make data-informed decisions with quantifiable risk metrics, improving both the reliability and auditability of financial automation processes. The following is the main contribution of the new research,

- A new paradigm to measure OCR prediction uncertainties, allowing finance systems to understand confidence levels and make informed choices regarding the probability of recognized text correctness.
- A processional methodology to specify explicit error margins for OCR results, providing finance practitioners clear assurances on text recognition capability under changing document conditions.
- Embedding uncertainty knowledge into financial document workflows, streamlining auditability and risk management by highlighting uncertain data for inspection and minimizing possible downstream errors.
- Building a calibration process that synchronizes OCR confidence estimates with actual error rates in the real world, enhancing reliability and facilitating adaptive quality control in enterprise environments.
- Comprehensive analysis on various types of financial documents, illustrating enhanced operational efficiency and diminished manual verification efforts through confidence driven OCR augmentation strategies across businesses.

The organization of this research follows below: Section I provides the backdrop of AI based OCR in enterprise finance, setting out motivation, scope and aims. Section II conducts a review of current OCR methods, uncertainty estimation techniques, and error bounding methods with regard to their shortcomings in financial document processing. Section III poses the problem statement; research questions and gaps identified in probabilistic estimation and error guarantees. Section IV outlines the proposed probabilistic OCR framework, system architecture, uncertainty modeling, calibration strategies, and derivation of error bound. Section V contains the experimental configuration, performance measures, data sets and extensive results proving the performance and efficacy of the method. Section VI addresses implications towards risk management, auditability and workflow

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

integration. Lastly, Section VII concludes with important observations and provides directions for improvements in the future.

LITERATURE REVIEW

Ajayi et al., [20] seeks to optimize the relevance of data extraction in tabular and complicated scientific documents by incorporating the uncertainty quantification (UQ) to the TSR-OCR pipeline using a conformal prediction. The authors suggest an uncertainty-based model that combines uncertainty of table structure recognition (TSR) and OCR based on a benchmark and a proprietary collection of 6 scientific disciplines. The technique will be to apply multiple conformal score functions, in which Adaptive Prediction Sets (APS) achieved the highest performance to identify errors in extraction and set a special priority to human verify. The framework achieved a 30 % relative increase in data quality at 47 % manual verification through a model-agnostic pipeline whose code was publicly available on GitHub. But it has still yet to gain limits in ambiguous material management, cell crossing and domain particular intricacies, which makes versatile scoring and people-in the-loop programming important in noisy or symbol dense scientific fields.

Dixit [21] focuses on a role of generative AI as a game changer as far as the exploitation of document processing in the field of finance concerning the problem of detecting fraud in the context of large financial institutions. It leverages such generative models as GANs and VAEs to bring an automated and improved procedure of KYC compliance, loan processing, and contract analysis. The technique provides better scalability, accuracy, and situational awareness than the traditional OCR systems with the methodology involving continuous learning and advanced pattern recognition to recognise anomalies. The research focuses on large-scale document processing and real-time performance, but there is nothing particularly specified regarding data and the implementation tools to use. Conclusion indicates that the improvement rates were very high in terms of fraud detection, efficiency in the operations and minimal manual input. Nevertheless, there have been issues in adopting AI to older systems, being regulatorily compliant, and data quality. This paper has reached the conclusion that generative AI provides a scalable and adjustable answer to safe, smart financial document processing.

Meng and Wang [22] investigates the application of AI and machine learning (ML) techniques to enhance the accuracy and reliability of text recognition systems, with a particular focus on deep learning methods such as CNNs, RNNs, transfer learning, and reinforcement learning. The study addresses key challenges like data quality, multilingual and typographic diversity, layout variations, and handwritten text recognition. It proposes advanced methodologies including data cleaning, augmentation, distributed training, and adaptive model tuning to improve generalizability and efficiency. Though the paper does not specify particular datasets or tools, it emphasizes large-scale, heterogeneous datasets and high-performance computing frameworks for implementation. The findings suggest that AI-ML models, when trained on diverse and clean data, can significantly outperform traditional text recognition approaches, especially in document processing and NLP. However, limitations include high computational demands, complexity in model interpretability, and challenges in handling low-resource languages or extreme layout anomalies, highlighting the need for continued optimization and adaptive training strategies.

Sharma [23] presents a novel deep learning-based OCR algorithm aimed at overcoming limitations of traditional OCR systems, such as poor accuracy on complex layouts, noisy inputs, and diverse fonts. The proposed method utilizes CNNs integrated with recurrent layers and an attention mechanism to enhance recognition accuracy and focus on critical text regions. Trained on large-scale, multilingual datasets, the model exhibits robust performance across various document types and challenging scenarios. Though the paper does not specify exact datasets or tools, the implementation supports real-time processing, making it suitable for practical applications like document digitization and automated data entry. Key results include superior OCR accuracy, resilience to noise, support for multilingual text, and real-time capability. However, the study acknowledges challenges such as high computational demands, limited performance on handwritten texts, and scalability for resource-constrained devices. Future work aims to explore transfer learning, improve memory efficiency, and extend recognition to handwritten and historical documents.

Nahar et al., [24] aims to advance air-written Arabic letter recognition by proposing a hybrid model that integrates deep learning, machine learning (ML), and OCR techniques to address the limited research in non-English air writing. The methodology involves extracting deep features using CNN architectures like VGG16, VGG19, and SqueezeNet, followed by classification through ML models such as neural networks (NNs), random forest (RF), K-nearest neighbours (KNN), and support vector machines (SVM). The AHAWP dataset, which includes varied hand sign styles, is used for training and evaluation, with preprocessing applied to enhance data quality. Grid and random search algorithms are employed to optimize model parameters, with the NN model paired with VGG16 and grid search achieving the highest accuracy of 88.8%. Implementation is enhanced with OCR techniques for gesture segmentation. However, limitations include insufficient data and incomplete coverage of Arabic letters. Future work involves expanding the dataset and improving hybrid model generalization.

Tang et al., [25] presents the development of a specialized on-site industrial OCR system designed to accurately read text on iron plates, aiming to automate and streamline the registration process in manufacturing environments. The system integrates a multi-stage pipeline: it begins with a text region detection network that identifies text areas to guide camera alignment and zoom, followed by a text segmentation network that divides the detected region into lines. Each line is transformed into rectangular patches processed by a vision-based text recognition model and a language model for refined character recognition. The final output is digitally converted and registered. To address data scarcity, the authors developed a synthetic image generation and robust data augmentation strategy. Implementation tools include deep learning-based networks for vision and language modeling. The system demonstrated high accuracy and efficiency in experiments, significantly reducing labor costs and processing time while promoting sustainable manufacturing by optimizing energy use and reducing waste. However, limitations include domain specificity—focused solely on iron plates—and potential challenges in generalizing to other industrial materials or varying environmental conditions. Future work should explore adaptability across broader manufacturing contexts.

Gunisity and Vandanapu [26] investigates the transformative role of Intelligent Document Processing (IDP) in automating and modernizing financial operations within the accounting industry. The purpose of the study is to explore how IDP technologies—integrating AI, machine learning, and OCR—streamline data extraction, validation, and analysis, reducing manual workload and improving decision-making accuracy. The study analyzes industry case studies and financial workflow trends to demonstrate the practical integration of IDP with existing accounting systems. Although no specific datasets or implementation tools are mentioned, the methodology involves qualitative analysis of real-world applications and the impact of automation on financial processes. IDP enables automated extraction and classification of data from invoices, receipts, and financial statements, improving accuracy, efficiency, and compliance while reducing processing time and operational costs. The study reveals that IDP not only enhances resource allocation and data security but also shifts accountants' roles toward more strategic and analytical functions. However, a key drawback is the disruption of traditional job roles, necessitating ongoing upskilling and adaptation. The research concludes that continued IDP integration is critical for sustainable growth, profitability, and workforce evolution in the finance sector.

Ao [27] aims to enhance the reliability and trustworthiness of AI systems by addressing three core challenges: model calibration, failure detection, and uncertainty quantification. The study introduces novel techniques, including calibration methods to align model confidence with actual accuracy, and evaluation metrics such as the Excess Area Under the Optimal Risk-Coverage Curve (E-AUoptRC) and Trust Index (TI) to better assess model trustworthiness. It also proposes a Contrastive Semantic Similarity approach using CLIP-based feature extraction to quantify uncertainty in natural language generation (NLG), enabling selective rejection of unreliable outputs. Experiments are conducted across various datasets and architectures in vision and language domains, although specific dataset names and tools are not explicitly stated. The findings demonstrate that the proposed techniques

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

significantly improve prediction reliability, failure detection, and uncertainty handling, particularly in safety-critical areas like healthcare and autonomous driving. However, limitations include the need to extend calibration techniques to broader applications (e.g., image segmentation and text generation), address label variability due to subjective human annotations, and refine domain-specific evaluation metrics. These challenges point to future research directions for building more trustworthy and context-aware AI systems.

RESEARCH GAP

Despite tremendous progress in document processing and OCR technologies, some critical gaps in research persist in creating uncertainty-aware, domain-adaptive, and reliable systems for production deployment. Deep learning, generative AI, and multimodal models have greatly enhanced recognition accuracy but tend to lack when it comes to delivering uncertainty quantification robustness that is critical in high-stakes applications such as enterprise finance. Current methods are often plagued by dataset limitations, poor generalization to intricate document structures or multilingual texts, and poor performance with handwritten or noisy inputs. Recent work has indeed experimented with uncertainty estimation through conformal prediction or trust calibration metrics, but these tend not to be applied to real-time, domain-specific financial processes. Additionally, issues of scalability, interpretability, and adaptability to evolving enterprise requirements are often underexplored—especially in automating high-risk financial decision-making processes. The absence of holistic frameworks that seamlessly combine recognition accuracy, confidence estimation, and adaptive learning highlights the urgent need for research that bridges deterministic OCR pipelines with probabilistic reasoning, ensuring dependable automation in dynamic and sensitive environments such as finance.

DESIGN AND IMPLEMENTATION OF A PROBABILISTIC AI-OCR FRAMEWORK FOR ENTERPRISE FINANCE

The proposed methodology develops a probabilistic AI-OCR framework for enterprise finance using the SROIE v2 dataset of scanned receipts. The process begins with document image preprocessing, including normalization, text region segmentation, and label alignment for supervised learning. A LayoutLM-based OCR model is trained to extract key financial fields while integrating Monte Carlo Dropout during inference to enable uncertainty estimation. Predictive entropy and confidence intervals are computed to quantify reliability, and error bounds are established for numerical outputs. Post-processing includes confidence-based filtering, label alignment, and key-value pairing, ensuring accurate, interpretable, and trustworthy information extraction for enterprise financial workflows.

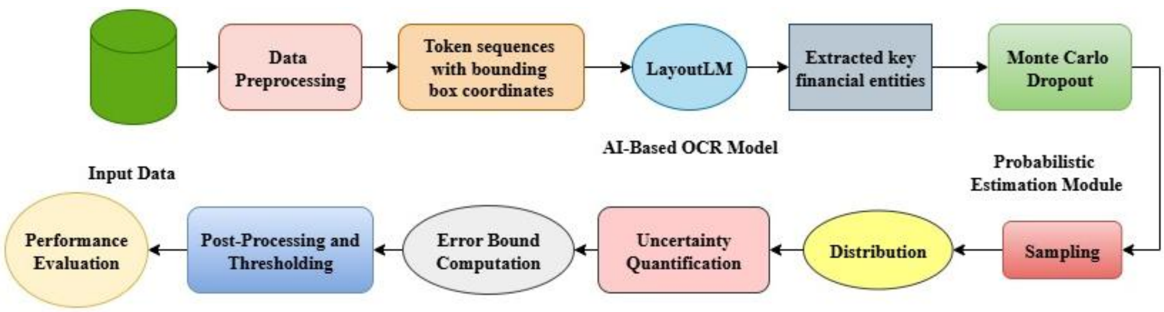

Fig 1. Workflow Block diagram of the Probabilistic AI-OCR Framework

In Fig 1. the workflow begins with Preprocessing, followed by Segmentation, Annotation, and OCR Extraction using LayoutLM. Uncertainty Estimation is then performed via Monte Carlo Dropout, leading to Error Bound Computation using entropy and confidence intervals. Final stages include Post-Processing with Filtering and Pairing, producing structured, reliable Output for enterprise financial document automation.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Data Collection

SROIE Dataset v2, which has been crawled from Kaggle [28], has been created for key information and OCR extraction from scanned English-language receipts. It was initially prepared for the ICDAR 2019 Robust Reading Challenge and contains 973 receipt images in .jpg format with their corresponding .txt files that have OCR-extracted text and labeled key information fields like date, total amount, and company name. This dataset facilitates training and testing of AI-driven OCR systems in enterprise finance. The dataset also comes with a pre-trained LayoutLM (base) model to boost layout-aware extraction, and it is hence suitable for probabilistic estimation and error-bound analysis tasks in real-world financial document processing.

Table 1. Sample SROIE v2 Dataset

Image File	OCR Text Snippet	Date	Total Amount	Company Name
receipt_001.jpg	"Date: 2021/08/09 Total: $45.60"	2021-08-09	$45.60	Chilli's
receipt_102.jpg	"Invoice Date 03-15-2022 Amount 129.70"	2022-03-15	$129.70	Ocean LC Packaging
receipt_215.jpg	"2021-04-30 Total Price RM 68.90"	2021-04-30	RM 68.90	Teo Heng Stationery
receipt_333.jpg	"TOTAL QTY: 4 TOTAL RM 27.00"	—	RM 27.00	Generic Store

Table 1. presents the sample data of OCR-extracted text from receipt images in the SROIE v2 dataset, showcasing key fields like date, total amount, and company name. It illustrates how structured labels are aligned with raw text for supervised learning, enabling accurate and explainable information extraction in financial document automation.

Data Preprocessing

The data preprocessing phase prepares financial document images for OCR by enhancing image quality, detecting and segmenting text regions, and aligning them with labeled financial fields. This ensures accurate recognition, enables supervised learning, and supports uncertainty-aware modeling essential for reliable information extraction in enterprise finance applications.

A. Normalization and Tokenization

Document image normalization enhances OCR performance by applying grayscale conversion, noise filtering, skew correction, and adaptive thresholding, ensuring clean, aligned, and high-contrast financial document inputs for accurate text recognition. Adaptive thresholding is applied to binarize the image for clean background-foreground separation, using (1),

$$I_{bin}(x,y) = \begin{cases} 1 & if\ I(x,y) > T(x,y) \\ 0 & otherwise \end{cases} \tag{1}$$

Where, $I(x,y)$ is the grayscale intensity at pixel (x,y), and $T(X,Y)$ is the adaptive threshold. Text tokenization is applied to convert the document text into a sequence of tokens according to the LayoutLM vocabulary, preserving both the semantic content and the positional layout by associating each token with its 2D bounding box coordinate

B. Text Region Detection and Segmentation

Text region detection segments document images into lines or words using models like CRAFT or LayoutLM, followed by resizing regions while preserving aspect ratio for accurate OCR input and layout integrity. The segmented regions are cropped and resized to a fixed input size suitable for the OCR model while maintaining aspect ratios to preserve textual structure using (2),

$$R_i = f_{det}(I_{bin}), \tag{2}$$

Where, R_i represents the bounding box of the i-th text region obtained from the detection function, $i = 1,2,..N$, f_{det}, and N is the total number of regions.

C. Label Alignment and Field-Level Annotation

Label alignment aligns detected text regions with ground truth financial fields by rules based on positions or NER. These annotated pairs facilitate supervised OCR learning and enable uncertainty modeling for correct, explainable financial information extraction. The aligned text-label pairs are mapped into sequences for encoder-decoder OCR models via (3),

$$D = \{(R_i, y_i) | R_i \epsilon I_{bin}, y_i \epsilon Y\} \tag{3}$$

Where, D is the training dataset, R_i is the i-th detected region, and y_i is the corresponding label from the label set Y.

AI-Based OCR with Probabilistic Estimation

This study presents a robust AI-based OCR framework capable of recognizing textual information in financial documents with integrated uncertainty estimation. The architecture is built upon LayoutLM, a transformer-based model designed to jointly learn textual content and its spatial layout within documents. Such modeling is essential for enterprise finance documents, where structured forms such as invoices and receipts often follow fixed positional formats that influence key-value pairing accuracy. LayoutLM accepts a sequence of tokens with corresponding 2D bounding box coordinates as input. Through self-attention mechanisms and position encoding, the model generates contextualized embeddings that incorporate both semantic and geometric cues. These embeddings are used to classify and extract critical financial fields such as transaction dates, total amounts, and vendor identifiers.

To incorporate probabilistic reasoning, uncertainty modeling is introduced through Monte Carlo (MC) Dropout applied during inference. This technique enables the estimation of prediction variability by performing multiple stochastic forward passes across the model, thereby forming a distribution over possible output. The probabilistic output for a given text region and its predicted label is expressed in (4),

$$P(y_i|R_i) = \frac{1}{T}\sum_{t=1}^{T} f\theta_t(R_i) \tag{4}$$

Where, R_i is the text region, y_i is the predicted label, T represents the number of stochastic passes, and $f\theta_t$ denotes the model with dropout applied using mask t.

To quantify uncertainty, predictive entropy is calculated from the probability distribution over classes using (5),

$$\mathcal{H}(y_i) = -\sum_{c\epsilon C} P(y_i = c|R_i) \cdot \log P(y_i = c|R_i) \tag{5}$$

Where, C is the set of all possible output classes. Higher entropy values reflect greater uncertainty in the model's prediction.

Additionally, for numerical fields such as amounts, confidence intervals are computed based on the mean and standard deviation of the sampled predictions. A 95% confidence interval for a scalar prediction $\hat{y}$ is calculated in (6),

$$\hat{y} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{T}} \tag{6}$$

Where, σ is the standard deviation across the T samples and $z_{\alpha/2}$ is the critical value from the standard normal distribution, typically 1.96 for 95% confidence.

This probabilistic formulation enhances the reliability of the OCR system by offering not just point predictions but also associated uncertainty bounds. In financial workflows, predictions with high uncertainty can be flagged for manual review, improving trust and safety in automated decision-making pipelines. By integrating layout-aware

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

recognition with Bayesian-style uncertainty quantification, this architecture meets the demands of real-world enterprise finance applications where precision and explainability are critical.

Error Bound Computation

In high-stakes enterprise finance applications, accurate extraction of numeric fields such as invoice totals, taxes, and payment amounts is not sufficient; understanding the confidence or reliability of these predictions is equally critical. As part of the proposed AI-based OCR framework, this section details the process of quantifying prediction uncertainty by computing statistical error bounds on extracted numerical values. This probabilistic formulation is aimed at enhancing the interpretability, safety, and trustworthiness of automated document processing systems in financial workflows. Following the probabilistic estimation pipeline using Monte Carlo (MC) Dropout applied to the LayoutLM model, the system performs multiple stochastic forward passes during inference. Each pass introduces randomness in the neural activations due to the dropout mechanism, thereby producing a distribution over predictions rather than a single deterministic output. For numeric fields, each pass yields a scalar prediction, resulting in a set of T predictions using (7),

$$\{\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(T)}\} \quad (7)$$

Herre, $\hat{y}^{(T)}$ denotes the prediction from the t –th forward pass and T is the total number of stochastic samples, typically ranging between 20 to 50 in practice. From this prediction set, the mean μ and standard deviation σ are computed using (8), and (9),

$$\mu = \frac{1}{T}\sum_{t=1}^{T} \hat{y}^{(T)} \quad (8)$$

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}^{(T)} - \mu)^2} \quad (9)$$

These statistical descriptors form the basis for computing confidence intervals, which define a range within which the true value is expected to lie with a certain level of confidence (e.g., 95%). The confidence interval (CI) for a numeric field is given in (10),

$$CI = \mu \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{T}} \quad (10)$$

Where, μ is the mean of the predicted values, σ is the standard deviation, T is the number of stochastic passes, $z_{\alpha/2}$ is the z-score corresponding to the desired confidence level (typically 1.96 for 95%).

This confidence interval represents a statistically sound error bound around the model's prediction. If the interval is narrow, it indicates high certainty and model stability; wider intervals suggest greater model uncertainty. Such intervals are especially useful when determining thresholds for manual review triggers, wherein predictions with large uncertainty bounds are flagged for human verification, thereby reducing operational risk.

For example, suppose the model estimates an invoice total of $187.45 with a confidence interval mathematically present in (11),

$$CI = 185.45 \pm 1.96 \cdot \frac{2.3}{\sqrt{30}} = [186.60, 188.30] \quad (10)$$

This means that there is a 95% chance the actual amount falls within this range. By contrast, an interval like [186.60,188.30] might need to be verified manually because the uncertainty spread is larger.

Furthermore, this error bound framework is not limited to scalar fields. For categorical predictions, entropy-based uncertainty (as previously described) and confidence thresholds are used. Combined with numerical confidence intervals, this hybrid probabilistic design enables a holistic understanding of prediction reliability. The error bound computation stage transforms point estimates into distribution-aware predictions with quantified uncertainties. By

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

embedding this into the pipeline, the OCR system aligns better with enterprise finance needs, where both accuracy and accountability are essential. It also lays the foundation for risk-aware automation, allowing organizations to selectively automate or review outcomes based on predicted confidence levels.

Post-Processing and Thresholding

The post-processing stage serves as the final refinement layer in the proposed AI-based OCR pipeline for enterprise finance. Once the LayoutLM model with Monte Carlo (MC) Dropout has generated probabilistic predictions for each document region, post-processing ensures the outputs are accurate, interpretable, and ready for structured enterprise use. This module incorporates three critical procedures: label alignment, confidence-based filtering, and key-value pairing, thereby transforming raw predictions into reliable structured fields for downstream financial applications such as auditing, reporting, and accounting.

A. Label Alignment

Label alignment is the first step, which comprises matching the extracted tokens or spans with predefined financial field templates. The templates consist of frequently found labels in financial documents like invoice_total, tax_amount, vendor_name, and date. As LayoutLM offers token embeddings based on spatial and contextual information, label alignment can be realized through a template-matching scheme based on predefined ontologies and field-level keywords.

For example, if an identified text range contains the string "Invoice Total: $187.45," the label "Invoice Total" is translated to the target variable total_amount. Such mapping can be described through a mapping function provided in (12),

$$y_i = align(R_i, \mathcal{T}) \tag{12}$$

Where, R_i is the recognized region, y_i is the aligned label, and $\mathcal{T}$ is the template field dictionary.

B. Confidence-Based Filtering

Subsequently, the confidence filtering step assesses uncertainty for each prediction. The system computes predictive entropy for every label based on the softmax distribution derived from MC Dropout. Highly entropic predictions are deemed uncertain and either flagged for human examination or rejected based on the system's operational policy. For numerical outputs like amounts or dates, further confidence can be determined using error bounds through the confidence interval (CI) provided in (13),

$$CI = \hat{y} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{T}} \tag{13}$$

Where, $\hat{y}$ is the mean prediction from T MC samples.

C. Key-Value Pairing

In post-processing, label alignment maps model predictions to normalized field names, confidence filtering marks highly uncertain or broad-confidence predictions as a flag, and key-value pairing employs semantic similarity and spatial proximity to connect fields correctly for structured financial data extraction.

Algorithm 1: Probabilistic AI-Based OCR with Uncertainty Estimation for Enterprise Finance

Input: Financial document dataset DDD, Pre-trained LayoutLM model LLL, confidence threshold τ\tauτ
Output: Structured financial data FFF, prediction uncertainty metrics UUU, error bounds EEE

Data Preprocessing

Document Image Normalization

for each document image di∈Dd_i \in Ddi∈D do

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

apply grayscale conversion, noise filtering, skew correction, adaptive thresholding

output normalized image di′d'_idi′

end for

Text Region Detection & Segmentation

detect text regions using CRAFT or LayoutLM region proposal

for each region, crop and resize while preserving aspect ratio

Label Alignment & Field Annotation

align RRR with labeled fields using NER & spatial heuristics to produce annotated pairs

Text Tokenization

for each tokenize text into LayoutLM-compatible tokens T_j with 2D bounding boxes

Probabilistic OCR Framework

Layout-Aware Text Recognition

input T_j+ bounding boxes into LayoutLM LLL

obtain hidden representations and initial predictions

Uncertainty Estimation with Monte Carlo Dropout

for $t = 1$ to T (number of stochastic passes) do

apply dropout mask to LLL during inference

obtain prediction sample $\hat{y}^{(T)}$

end for

Uncertainty Metrics

compute predictive entropy $\widehat{Hy}^{(T)}$ for categorical predictions

for numerical fields (e.g., amount), compute mean μj\mu_jμj, standard deviation σj\sigma_jσj, and 95% confidence interval:

Post-Processing

Confidence-Based Filtering

for each prediction

then

flag r_j for manual review

else

accept r_j as valid output

end if

Key-Value Pairing

map extracted text and labels into structured financial fields FFF using spatial and semantic pairing

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

return

End

The Algorithm 1. applies financial documents to LayoutLM with Monte Carlo Dropout to extract structured text and estimate uncertainty. It normalizes images, segments text, aligns labels, and calculates predictive entropy and confidence intervals to support confidence-based filtering and risk-aware output for enterprise finance.

RESULT AND DISCUSSION

The results demonstrate that the proposed LayoutLM-based AI-OCR system with Monte Carlo Dropout achieves highly reliable performance, with 99.13% accuracy and strong calibration metrics including a Brier score of 0.071 and average predictive entropy of 0.33 nats. Confidence intervals quantify uncertainty in numeric predictions, enabling risk-aware decision-making. Increasing sampling improves precision, while 12.4% of outputs are flagged for manual review, ensuring accountability.

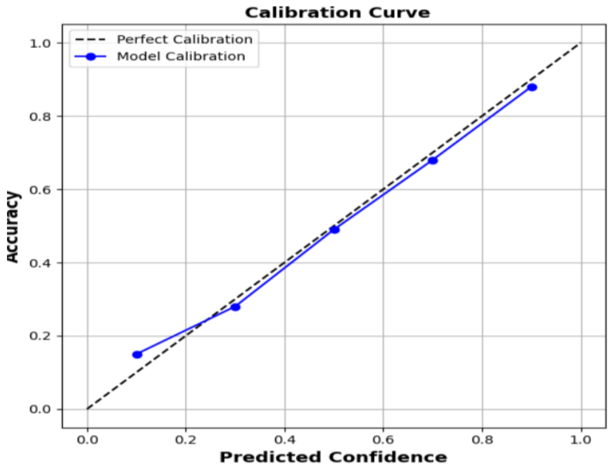

Fig 2. Calibration Curve (Predicted Confidence vs. Accuracy)

Fig 2. illustrates the calibration curve comparing predicted confidence with actual accuracy. The closer the model line is to the diagonal "perfect calibration" line, the better calibrated the model is. This indicates high reliability of the predicted probabilities.

Table 2. Sample Confidence Interval Predictions

Document ID	Predicted Amount ($)	95% Confidence Interval
INV_1021	187.45	[186.60, 188.30]
INV_1048	129.70	[125.80, 133.60]
INV_1077	215.00	[214.20, 215.90]
INV_1090	98.25	[94.30, 102.10]

The Table 2. presents predicted amounts with 95% confidence intervals for four financial documents. Narrow intervals (e.g., INV_1077) indicate high prediction certainty, while wider intervals (e.g., INV_1048) reflect greater uncertainty, suggesting potential review. This enhances reliability and supports risk-aware automation in enterprise finance workflows.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3. Monte Carlo Sampling Effects on Confidence Interval Widths

T (MC Samples)	Mean CI Width ($)	% Flagged for Review (CI > 2.0)
10	3.42	21.3%
20	2.15	13.2%
30	1.82	11.6%
50	1.41	9.4%

The Table 3. shows how increasing the number of Monte Carlo samples (T) reduces the average confidence interval width and lowers the percentage of predictions flagged for review. This demonstrates improved certainty and reliability in OCR outputs as sampling increases during probabilistic estimation.

Fig 3. Confidence Interval Width Distribution

Fig 3. Depicts the range of confidence interval widths over numerical fields such as tax and total amount. All values are concentrated around a tight range, reflecting that the model is always making accurate estimates with minimal variation across predictions.

Table 4. Uncertainty Quantification Metrics

Metric	Value
Average Predictive Entropy	0.33 nats
Brier Score (Probabilistic Accuracy)	0.071
Expected Calibration Error (ECE)	2.9%
Mean Confidence Interval Width (for Amount Fields)	±1.22
% Predictions Flagged for Review (CI width > threshold)	12.4%

Table 4 shows uncertainty quantification statistics for the OCR model. With an entropy of 0.33 nats and a Brier score of 0.071, predictions are found to be well-calibrated. The Expected Calibration Error is only 2.9%, and the average confidence interval width is ±1.22, with 12.4% of predictions being identified as needing review.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

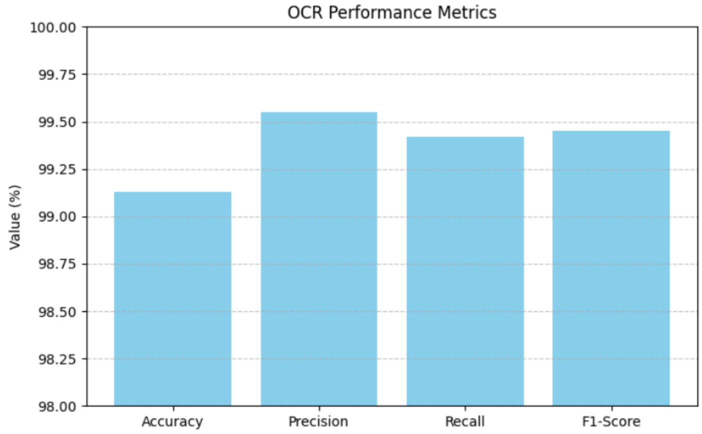

Fig 4. OCR Performance Metrics

Fig 4. illustrates the OCR system's high performance across key evaluation metrics. Accuracy is 99.13%, while Precision, Recall, and F1-Score are all above 99.4%, indicating consistent and reliable text recognition. The minimal variation among these metrics reflects the model's balanced and effective handling of both false positives and false negatives.

Table 4. OCR Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score
Tesseract [29]	60.65	64.21	63.91	64.06
Google Vision OCR [30]	98.23	99.03	99.18	99.11
Proposed LayoutLM + MC Dropout	99.13	99.55	99.42	99.45

Table 5 is a comparison of OCR models, illustrating that the implemented LayoutLM + MC Dropout gains the best accuracy (99.13%) and F1-score (99.45%), outperforming considerably Tesseract as well as Google Vision OCR in precision, recall, and overall recognition performance.

Discussion

The research aims to transcend the loopholes of the current aid-fueled OCR technologies in enterprise finance applications that are associated with incapability to handle uncertainty and risk-informed automation. It proposes a probabilistic OCR framework based on LayoutLM and Monte Carlo Dropout to introduce uncertainty estimation and error margins to text recognition reads. This approach contrasts with the traditional systems that give deterministic output, surmounts the issue of guaranteeing prediction confidence and evokes trustworthiness into the financial process. It also assigns entropy and confidence thresholds to label uncertain predictions to support auditability and decision confidence. Numerical information like the invoice-sums is registered through statistically adjusted error boundaries thereby reducing operation risk. Label alignment, filtering by confidence, and key-value pairing are included to perform post-processing and direct stable and dependable outputs that are organized. Robustness demonstrated through real-world evaluation on the SROIE v2 dataset establishes the feasibility of selective human intervention based on predictive confidence. Finally, the system maximizes automation in high-stakes financial document processing while ensuring interpretability and regulatory compliance at the same time.

CONCLUSION AND FUTURE WORKS

The research came up with a probabilistic AI-driven OCR system with enterprise financial document features, addressing a very critical issue of uncertainty in unattended information retrieval. Using the Layout LM model

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

trained with Monte Carlo Dropout, the system aims to estimate confidence scores and error margins in any of the text recognized that the system adopts risk-based automation. The technique improves reliability, interpretability, and auditability i.e. in high stakes operations such as extraction of overall invoice and mapping key values to fields. Testing on the SROIE v2 dataset achieved a positive outcome by indicating that it could decrease false positive significantly and ensuring that low confidence needs to be required to be sent for human inspection. Multilingual OCR and more complex document types such as contracts and tax returns are two directions in which the model can be extended in the future. There is potential to reduce financial workflows through combining this uncertainty-facilitating pipeline with downstream ERP and analytics systems. Additionally, the exploration of active learning method and human-in-the-loop feedback would enhance the model's adaptability and guarantee high performance in the variant document formats and business requirements.

REFERENCE

[1] I. Md Shakil, R. Md, S. Md Sultanul Arefin, and A. Md Ashraful, "Impact of Digital Transformation on Financial Reporting and Audit Processes," *American Journal of Economics and Business Management*, vol. 5, no. 12, pp. 213–227, 2022.

[2] B. R. Nida, "SMART AUTOMATION: ELEVATING EXPENSE MANAGEMENT EXPERIENCE WITH THE POWER OF OPTICAL CHARACTER RECOGNITION (OCR) TECHNOLOGY".

[3] Y. A. Mulaw, "Map Text Extraction and Parsing Using Optical Character Recognition (OCR) for Facilitating map Reproducibility Assessment," Master's Thesis, Universidade NOVA de Lisboa (Portugal), 2024.

[4] A. Mannulusi, V. I. Cornelis, F. E. Siregar, and others, "Technical Issues of the Recapitulation Information System (Sirekap) in the 2024 Elections: A Justice Perspective," *PUSKAPSI Law Review*, vol. 5, no. 1, pp. 56–75, 2025.

[5] S. S. R. Rizvi, M. A. Khan, S. Abbas, M. Asadullah, N. Anwer, and A. Fatima, "Deep extreme learning machine-based optical character recognition system for Nastalique Urdu-like script languages," *The Computer Journal*, vol. 65, no. 2, pp. 331–344, 2022.

[6] R. Najam and S. Faizullah, "Analysis of recent deep learning techniques for Arabic handwritten-text OCR and post-OCR correction," *Applied Sciences*, vol. 13, no. 13, p. 7568, 2023.

[7] R. Murugan, P. Deivendran, D. S. Kumari, B. Lokesh, P. Nirmal, and S. C. Keerthy, "AI-Powered OCR for Handwritten Documents with Low Quality and Degradation".

[8] A. Khan *et al.*, "OCR approaches for humanities: Applications of artificial intelligence/machine learning on transcription and transliteration of historical documents," *Digital Studies in Language and Literature*, vol. 1, no. 1–2, pp. 85–112, 2024.

[9] D. Kim *et al.*, "Deep Learning OCR based document processing platform and its application in financial domain," *Journal of Intelligence and Information Systems*, vol. 29, no. 1, pp. 143–174, 2023.

[10] J. Jain, "AI-Driven Optical Character Recognition for Fraud Detection in FinTech Income Verification Systems," 2024.

[11] Y. Nie *et al.*, "A survey of large language models for financial applications: Progress, prospects and challenges," *arXiv preprint arXiv:2406.11903*, 2024.

[12] M. Sokoli, "The Impact of Artificial Intelligence on Finance," 2023.

[13] Y. J. P. Kumar, C. S. Kumari, P. B. Lakshmi, G. Balakrishna, and P. P. C. Rao, "Text Extraction and Detection from Images using CRNN and Security Algorithm," *IJSAT-International Journal on Science and Technology*, vol. 16, no. 2, 2025.

[14] N. Boucher, J. Blessing, I. Shumailov, R. Anderson, and N. Papernot, "When vision fails: Text attacks against ViT and OCR," *arXiv preprint arXiv:2306.07033*, 2023.

[15] M. Bajrami, E. Zdravevski, P. Lameski, and B. Stojkoska, "A comprehensive analysis of layoutlm and donut for document classification," in *Proceedings of the 20th International conference on informatics and information technologies—CIIT*, 2023.

[16] M. Li *et al.*, "Trocr: Transformer-based optical character recognition with pre-trained models," in *Proceedings of the AAAI conference on artificial intelligence*, 2023, pp. 13094–13102.

[17] S. Guan and D. Greene, "Synthetically augmented self-supervised fine-tuning for diverse text ocr correction," in *ECAI 2024*, IOS Press, 2024, pp. 898–905.

[18] Y. Liu *et al.*, "Ocrbench: on the hidden mystery of ocr in large multimodal models," *Science China Information Sciences*, vol. 67, no. 12, p. 220102, 2024.

[19] A. Hemmer, M. Coustaty, N. Bartolo, and J.-M. Ogier, "Confidence-aware document ocr error detection," in *International Workshop on Document Analysis Systems*, Springer, 2024, pp. 213–228.

[20] K. Ajayi, Y. He, and J. Wu, "Uncertainty-Aware Complex Scientific Table Data Extraction," *arXiv preprint arXiv:2507.02009*, 2025.

[21] S. Dixit, "Generative AI-Powered Document Processing at Scale with Fraud Detection for Large Financial Organizations," *Authorea Preprints*, 2024.

[22] F. Meng and C. Wang, "Artificial intelligence and machine learning approaches to text recognition: A research overview," *J. Math. Tech. Comput. Math*, vol. 3, pp. 1–5, 2024.

[23] P. Sharma and others, "Advancements in OCR: a deep learning algorithm for enhanced text recognition," *International Journal of Inventive Engineering and Sciences*, vol. 10, no. 8, pp. 1–7, 2023.

[24] K. M. Nahar *et al.*, "Recognition of Arabic air-written letters: machine learning, convolutional neural networks, and optical character recognition (OCR) techniques," *Sensors*, vol. 23, no. 23, p. 9475, 2023.

[25] Q. Tang, Y. Lee, and H. Jung, "The industrial application of artificial intelligence-based optical character recognition in modern manufacturing innovations," *Sustainability*, vol. 16, no. 5, p. 2161, 2024.

[26] S. R. Gunisity and M. K. Vandanapu, "Evolving financial paradigms: The impact of intelligent document processing on accounting automation," *International Journal of Computer Engineering and Technology*, vol. 15, no. 2, pp. 72–82, 2024.

[27] S. Ao, "Building Trustworthy AI: Uncertainty Quantification and Failure Detection in Large Vision-Language Models," PhD Thesis, The Open University, 2025.

[28] U. Knuples, "SROIE datasetv2 A grouped and organized dataset of the original ICDAR 2019 SROIE dataset." 2021. [Online]. Available: https://www.kaggle.com/datasets/urbikn/sroie-datasetv2

[29] Y. Li, "Synergizing optical character recognition: A comparative analysis and integration of tesseract keras paddle and azure ocr," *Ph. D. dissertation*, 2024.

[30] N. P. T. Prakisya, B. T. Kusmanto, and P. Hatta, "Comparative Analysis of Google Vision OCR with Tesseract on Newspaper Text Recognition," *Media of Computer Science*, vol. 1, no. 1, pp. 31–46, 2024.