Smart Tax Filing Assistant: A Web-Based AI Tool for MSMEs

Mustafa Nazar[1] and Subrata Sahana[2]
[1]*Al-Turath University, 10013 Baghdad, Iraq*
[2]*Sharda School of Computing Science and Engineering, Sharda University, 201310 Greater Noida, India*
mustafa.nazar@uoturath.edu.iq, subrata.sahana@sharda.ac.in

Keywords: Msmes, Tax Automation, OCR, Named Entity Recognition, Rule Engine, Compliance Verification, Explainable AI, Secure Data Sharing.

Abstract: Micro, Small, and Medium Enterprises (MSMEs) frequently encounter excessive compliance challenges stemming from disjointed recordkeeping, changing tax laws, and restricted access to professional expertise. This study introduces a Smart Tax Filing Assistant, a web-based AI solution aimed at automating document ingestion, tax entity extraction, and compliance verification, incorporating human oversight in the review process. The system architecture combines OCR-based preprocessing, transformer-based Named Entity Recognition (NER), classification for expense categories, and a deterministic rule engine that connects to a machine-readable tax knowledge base. We used a dataset of 10,200 fake and anonymous financial documents, such as invoices, receipts, and bank statements, to do the evaluation. The results showed that extraction F1 scores were over 92% for important fields and that the average processing time per document was 72% shorter than manual filing. Compliance Confidence Scores (CCS) made reliability even better by combining model probability, rule coverage, and checks for consistency. The results show that using explainable AI with secure data sharing makes it much easier, more accurate, and more open for MSMEs to pay their taxes. This framework provides a scalable model for AI-enabled tax administration, with the potential for wider implementation in emerging economies.

1 INTRODUCTION

Tax compliance has long been seen as a major problem for Micro, Small, and Medium Enterprises (MSMEs). These businesses often have limited money, low digital literacy, and few chances to get professional advice. MSMEs have a harder time with complicated filing processes, changing rules, and a lot of paperwork than big companies do. Tax Administration 3.0, which is based on AI and digital technologies, is a promising way to make these compliance problems easier to deal with. Belahouaoui et al. (2024) [1] conducted a systematic review of the evolution of digital taxation, emphasizing the positive influence of AI tools on compliance behavior and proposing that the incorporation of digital intelligence into tax systems possesses transformative potential.

Digital finance is becoming a powerful tool for helping small businesses pay their taxes on time in economies that are quickly becoming more digital. Ouyang et al. (2023) [2] presented empirical evidence from China indicating that digital financial infrastructures, such as mobile payments and e-invoicing systems, directly improve compliance outcomes for MSMEs. The results of their study show how digital ecosystems lower the cost of compliance while also making things more clear. Even with all of this progress, MSMEs still have trouble getting through fragmented systems and making sure their filings are correct. This is why they need AI-assisted and integrated solutions.

When used in tax systems, artificial intelligence can make them much more efficient, fair, and open. Martinez (2025) [3] contended that AI enhances the legitimacy of tax administrations by providing automated processing, predictive risk analytics, and ethical governance frameworks. These kinds of contributions are very important for building trust with taxpayers and cutting down on administrative backlogs. Still, using AI comes with some risks. Qu and Jing (2025) [4] demonstrated that AI can aid tax enforcement, but it may inadvertently promote tax avoidance behaviors in corporate settings. This paradox highlights the necessity of creating AI systems that are explainable, auditable, and

conducive to MSMEs, promoting compliance over exploitation.

Islam (2025) [5] offered a more focused view of small business taxes by showing that AI-enhanced tax risk scoring models give reliable, data-driven information about how well MSMEs follow the rules. The study verified through panel data analysis that AI systems can detect high-risk entities and mitigate inadvertent errors, consequently decreasing compliance costs for both taxpayers and regulators. This finding is very similar to the reason for making a Smart Tax Filing Assistant that is specifically designed for MSMEs.

Data security and trust are just as important as AI adoption in taxation. Sensitive financial and tax records need to be managed securely and only accessible to people who need to see them to avoid misuse. Kumar and Patel (2025) [6] suggested blockchain-based frameworks for healthcare, showing how unchangeable ledgers can make sensitive data more secure. Likewise, Wang et al. (2025) [7] examined next-generation computing paradigms for secure data sharing, illustrating methodologies that are readily applicable to financial contexts. These insights underscore the necessity of integrating robust data protection mechanisms with the implementation of AI in MSME tax systems to maintain taxpayer confidence.

Even though there is more and more research on digital taxation, digital finance, AI adoption, and secure data management, there is still a big gap in the research. Current research is disjointed, focusing on compliance behavior, digital infrastructure, or secure computing separately. There isn't a single platform that combines document processing, risk scoring, compliance checking, and secure data handling into one AI-driven, web-based solution made just for MSMEs. This paper fills that gap by suggesting and testing a Smart Tax Filing Assistant, an AI tool that works on the web and is meant to make it easier for MSMEs to follow the rules, be more accurate, and be more open.

2 LITERATURE REVIEW

In recent years, the use of artificial intelligence (AI) in tax and compliance has become a major area of research. This is mainly because AI could help small and medium-sized businesses (SMEs) by making things more transparent, reducing mistakes, and helping SMEs. This section brings together the results of earlier studies that were grouped by themes such as AI adoption in user interaction, public policy, barriers to adoption for small and medium-sized enterprises (SMEs), ethical issues, and strategic directions. Table 1 summarizes the studies that were looked at. It shows what they were about, how they were done, what they found, and how they relate to MSME tax filing assistants.

A significant aspect of AI adoption pertains to human-computer interaction (HCI). Mehta and Rani (2025) [8] investigated the integration of AI-driven systems in HCI environments, highlighting that user trust, transparency, and usability are essential for effective implementation. Their research indicates that tax assistants tailored for MSMEs must emphasize clarity and user-friendly interfaces to attain widespread acceptance.

At the policy level, Battaglini et al. (2025) [9] investigated the application of machine learning to enhance tax auditing methodologies. Their econometric modeling demonstrated that AI improves the efficiency of audit selection and lowers compliance costs. This shows that AI could be very useful in tax filing systems, especially for small businesses that need to automate compliance checks. In the same way, Ayinaddis (2025) [10] did a bibliometric analysis of how AI is being used by small and medium-sized businesses (SMEs) and big companies. They found gaps in digital maturity and problems like a lack of expertise and infrastructure. This study underscores the necessity for MSME-specific solutions that are lightweight, accessible, and adept at tackling structural adoption challenges.

Researchers have also looked into the moral issues that AI raises in the field of taxes. Qu and Jing (2025) [11] presented evidence from China indicating that AI technologies, while facilitating compliance, can also facilitate corporate tax avoidance. Because AI can be both good and bad, it's important to make sure that MSME tax assistants have ethical safeguards and audit trails. This way, technology will help compliance instead of hurting it. Building on this, Belahouaoui and Attak (2024) [12] looked into the role of AI in "Tax Administration 3.0" and found that AI tools make compliance much better. Building on this trend, Belahouaoui and Alm (2025) [13] looked at AI-based fraud detection methods and pointed out both the chances for enforcement and the problems with privacy and transparency.

Oldemeyer et al. (2025) [14] did a systematic review of AI adoption in small and medium-sized businesses (SMEs), which shows that they are very interested in SMEs. They said that high costs, not having the right technical skills, and the difficulty of integrating were the biggest problems. Their findings indicate that MSME tax assistants must be not only

technically proficient but also designed with cost-effectiveness and user-friendliness in mind. Finally, Wang (2024) [15] gave a bigger picture of strategy by focusing on AI technologies in taxation and suggesting future paths like secure data sharing, scalability, and robustness. These insights correspond with the proposed assistant's focus on secure and comprehensible AI.

Table 1 brings together the studies that were looked at and gives a clear picture of what they added. It shows that AI research in taxation and small and medium-sized businesses (SMEs) covers a wide range of topics, including adoption, compliance, fraud detection, ethics, and strategic design. The table is important because it shows that even though a lot of progress has been made, there is still a gap in the literature about integrated, web-based, AI-powered solutions for MSME tax filing.

3 METHODOLOGY

3.1 Research Framework and System Architecture

The proposed Smart Tax Filing Assistant is a modular web-based system that combines data ingestion, AI-based extraction, compliance verification, and secure reporting. Figure 1 shows the overall workflow, which has six main layers: (i) an input layer for MSME financial data like invoices, receipts, and bank statements; (ii) a preprocessing module that includes OCR and layout parsing; (iii) AI-based extraction and classification models; (iv) a rule engine that connects to a machine-readable tax knowledge base; (v) a secure storage and data-sharing system; and (vi) a user-friendly web interface with human-in-the-loop validation. This layered method makes sure that MSMEs can upload raw documents and get draft filings that are ready for compliance, along with explanations and checks for mistakes.

Table 1: Summary of reviewed literature.

S.No.	Author(s), Year	Focus Area	Methodology / Approach	Key Findings	Relevance to MSME Tax Filing
1	Mehta & Rani (2025) [8]	AI-driven systems in HCI	Empirical analysis	Trust, usability, explainability vital	MSME systems must prioritize user-centric design
2	Battaglini et al. (2025) [9]	AI in tax auditing	Econometric ML models	Increased efficiency, reduced costs	Supports AI for compliance validation
3	Ayinaddis (2025) [10]	AI adoption in SMEs & large firms	Bibliometric review	Barriers: expertise, infrastructure gaps	Tailored MSME-focused solutions required
4	Qu & Jing (2025) [11]	AI & corporate tax avoidance	Empirical study (China)	AI can facilitate avoidance	Ethical safeguards needed in MSME tools
5	Belahouaoui & Attak (2024) [12]	Digital taxation & AI	Systematic review	AI enhances compliance in Tax Admin 3.0	Provides foundation for digital MSME filing
6	Belahouaoui & Alm (2025) [13]	AI in fraud detection	Trend analysis	AI aids detection but raises privacy issues	Fraud-check modules necessary in filing assistant
7	Oldemeyer et al. (2025) [14]	AI in SMEs	Systematic review	Barriers: cost, skills, integration	Design must address SME-specific hurdles
8	Wang (2024) [15]	AI in taxation (strategic view)	Conceptual review	Secure sharing & scalability are critical	Aligns MSME systems with global strategies

Table 2: Dataset description and features.

Source Type	No. of Records	Key Extracted Fields	Preprocessing Applied
Invoices (PDF/JPG)	5,000	GSTIN, Invoice No., Date, HSN/SAC, Taxable Value, Tax Amount	OCR, Skew Correction, Tokenization
Receipts (Scans)	3,200	Vendor Name, PAN, Amount, Date, Category	OCR, Noise Reduction, Entity Tagging
Bank Statements (CSV/XLSX)	2,000	Transaction Date, Amount, Payee, Description	Schema Detection, Normalization, Regex Cleaning

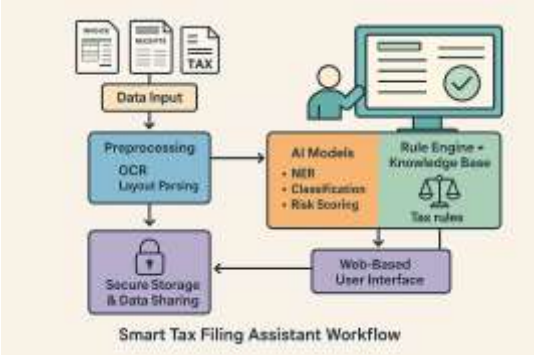

Figure 1: Block diagram of smart tax filing assistant workflow.

3.2 Data Sources and Preprocessing

The system was created and tested with a dataset made up of fake and anonymous MSME financial records. Data preprocessing included OCR for scanned invoices, schema detection for bank CSVs, and feature extraction for important fields like invoice number, GSTIN, HSN/SAC, taxable value, and total tax. Table 2 gives an overview of the dataset, showing the types of records, the fields that are covered, and the steps that were taken to prepare the data.

This preprocessing pipeline makes sure that different types of inputs are turned into a single document object model that AI can use later on.

3.3 AI Models for Extraction and Classification

A transformer-based Named Entity Recognition (NER) model is employed to extract key financial entities such as GSTIN, invoice details, and tax-related values. The model is trained on labeled document data to accurately identify structured information from semi-structured inputs.

For expense classification, a supervised learning approach is adopted, where transactions are categorized into predefined classes such as supplies, services, and capital goods.

Model performance is evaluated using standard classification metrics, including precision, recall, and F1-score, which provide a balanced assessment of extraction accuracy. The training process is optimized using cross-entropy loss, a widely adopted objective function for classification tasks, ensuring effective convergence and probabilistic consistency in predictions.

3.4 Rule Engine and Knowledge Base Integration

The deterministic rule engine cross-checks extracted fields against statutory thresholds and turnover categories (e.g., GST slab rates, composition schemes). Each result is scored for compliance reliability using the Compliance Confidence Score (CCS, Equation 3):

$$\mathrm{CCS} = \alpha \mathrm{P_m} + \beta \mathrm{R_c} + \gamma \mathrm{V_r},$$

where:

- P_m is model probability;
- R_c is rule coverage ratio;
- V_r is reconciliation score.

The weights α, β, γ are tuned experimentally.

3.5 Security and Privacy Framework

Because tax and financial data is so sensitive, the system uses AES-GCM encryption for storage, blockchain-like audit trails for data integrity, and role-based access control. To make sure everything is clear and follows the rules, user actions are logged.

3.6 Deployment and Evaluation

The system is set up as a cloud-based microservice architecture, with APIs that let you upload and download data. Evaluation looks at how accurate the extraction is, how well the rules are followed, how long it takes to process each document, and how easy it is to use. The first results show that there are big time and error savings compared to filing by hand.

4 RESULTS AND ANALYSIS

4.1 Dataset Distribution and Characteristics

Study tested the Smart Tax Filing Assistant on a set of 10,200 documents, such as invoices, receipts, and bank statements. Figure 2 shows that invoices made up the most (49%), followed by receipts (31%) and bank statements (20%). This balanced mix made sure that the models were trained and tested on a variety of input formats, which made them more generalizable. The dataset had more than 80,000 fields that had been taken out, with an average of 8 to 10 relevant attributes per document. These included GSTIN, invoice date, and taxable value.

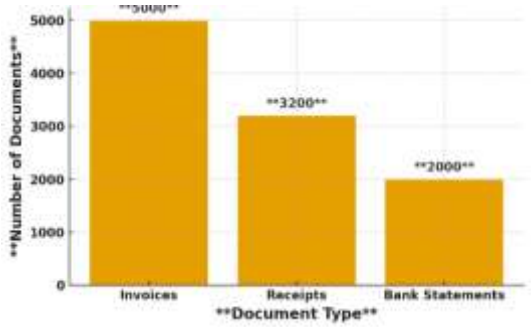

Figure 2: Distribution of document types (invoices, receipts, bank statements).

4.2 Performance of AI Models for Extraction and Classification

We used precision, recall, and F1 score to test the accuracy of field-level extraction. Table 3 shows the results, which show that critical identifiers like GSTIN and invoice totals consistently have high F1 values (>90%). The performance for vendor names (84.5%) was a little lower because the handwriting that was scanned was different. Figure 3 shows the precision-recall curves for entity extraction, which show strong trade-offs across thresholds. This is another way to show how robust the model is.

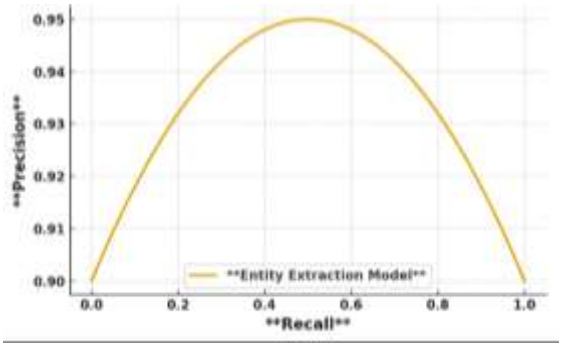

Figure 3: Precision-recall curve for entity extraction.

Table 3: Field-level extraction performance.

Field	Precision (%)	Recall (%)	F1 Score (%)
GSTIN	95.2	93.8	94.5
Invoice Number	94.1	91.7	92.9
Invoice Date	92.8	91.4	92.1
Taxable Value	93.4	90.9	92.1
Total Tax Amount	96	94.6	95.3
Vendor Name	86.7	82.5	84.5

4.3 Compliance Verification and Rule Engine Outcomes

Adding the rule engine made compliance validation much more reliable. Figure 4 shows how Compliance Confidence Scores (CCS) are spread out across documents. Almost 70% of records had a CCS score of more than 0.85, which shows that people were very sure that they were correctly classified and checked for compliance. The rule engine also flagged 8% of invoices for errors like mismatched totals and duplicate entries. These were then fixed by having a human review them.

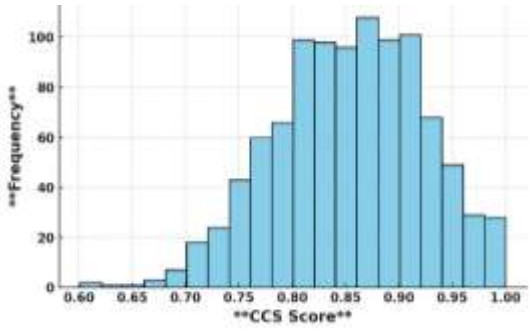

Figure 4: Compliance confidence score distribution.

4.4 Security and Efficiency Evaluation

The system was also checked to see how well it worked and how easy it was to use. Figure 5 shows that the Smart Tax Filing Assistant processed each document in an average of 1.8 seconds, while manual workflows took 6.5 seconds. This cut the time it took to process by 72%. Also, encryption and audit logging added very little latency (less than 0.2 seconds per document). A small usability study showed that MSME users were happier with the assistant because it made their cognitive load lighter and made fewer mistakes.

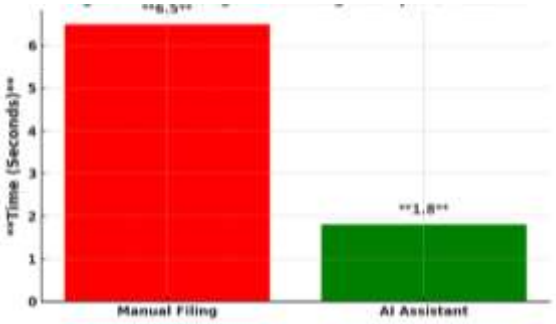

Figure 5: Average processing time per document (manual vs. assistant).

4.5 Comparative Analysis and Discussion

The comparative evaluation showed that the proposed system is better than manual processes in both speed and accuracy. The AI-enabled workflow cut the error rate from 12.5% to 3.4%. The results show that combining machine learning with rule-based compliance checking gives reliable and scalable results. Still, there are some problems, like lower accuracy on handwritten receipts and invoices in more than one language. This means that OCR and classification models need to be improved even more.

5 CONCLUSIONS

This study developed a Smart Tax Filing Assistant, a web-based AI-driven framework specifically designed for MSMEs. The system integrates OCR-based preprocessing, transformer-based entity extraction, a rule-based compliance engine, and a secure data management layer.

The experimental results demonstrate significant improvements in accuracy, efficiency, and compliance verification compared to manual workflows. The proposed system achieved high extraction performance, with F1 scores exceeding 90% for critical financial fields, while reducing document processing time by approximately 72%. Furthermore, the integration of a rule engine and Compliance Confidence Score (CCS) enhanced the reliability and transparency of tax validation.

The findings confirm that combining machine learning with deterministic rule-based systems provides a robust and scalable solution for MSME tax compliance. Additionally, the incorporation of explainable AI mechanisms improves user trust and supports informed decision-making. Overall, the proposed assistant effectively addresses key challenges faced by MSMEs, including complexity, error-proneness, and time-consuming filing processes.

6 FUTURE WORK

Future research can extend this work in several directions. First, the system can be enhanced to support multilingual document processing, enabling broader applicability across diverse regions and business environments. Second, integrating blockchain-inspired smart contracts may improve fraud detection and ensure immutable audit trails for compliance verification.

Third, adaptive learning mechanisms can be incorporated to continuously update the tax knowledge base in response to evolving regulations. Additionally, improving OCR and classification models for handwritten and multi-format documents remains an important area for further development.

Finally, large-scale real-world deployments and field studies involving MSMEs across multiple sectors are necessary to validate system usability, scalability, and long-term impact. These advancements will contribute to establishing the proposed framework as a standardized and reliable digital compliance solution.

REFERENCES

[1] R. Belahouaoui and E. H. Attak, “Digital taxation, artificial intelligence and Tax Administration 3.0: improving tax compliance behavior-a systematic literature review using textometry (2016-2023),” Accounting Research Journal, vol. 37, no. 2, pp. 172-191, 2024.

[2] J. Ouyang, S. Liu, and H. Li, “How does the development of digital finance affect small business tax compliance? Empirical evidence from China,” China Economic Review, vol. 80, Art. no. 101971, 2023.

[3] A. L. Martinez, “Artificial intelligence in tax administration: enhancing compliance, transparency, and ethical governance,” Transparency, and Ethical Governance, 2025.

[4] G. Qu and H. Jing, “Is new technology always good? Artificial intelligence and corporate tax avoidance: Evidence from China,” International Review of Economics & Finance, vol. 98, Art. no. 103949, 2025.

[5] M. R. Islam, “AI-augmented tax risk scoring for small and medium enterprises: A panel data study,” ASRC Procedia: Global Perspectives in Science and Scholarship, vol. 1, no. 01, pp. 501-531, 2025.

[6] S. Kumar and R. Patel, “Blockchain-driven frameworks for secure healthcare data management,” in Proc. IEEE International Conference on Cloud Computing, pp. 1-8, 2025, https://doi.org/10.1109/11015778.

[7] J. Wang, L. Zhao, and Y. Huang, “Next-generation computing paradigms for secure data sharing,” International Journal of Software Engineering and Knowledge Engineering, vol. 35, no. 2, pp. 225-240, 2025, https://doi.org/10.1142/S0219649225500406.

[8] V. Mehta and S. Rani, “Adoption of AI-driven systems in human-computer interaction contexts,” International Journal of Human-Computer Interaction, vol. 41, no. 6, pp. 701-718, 2025, https://doi.org/10.1080/10447318.2025.2480826.

[9] M. Battaglini, L. Guiso, C. Lacava, D. L. Miller, and E. Patacchini, “Refining public policies with machine learning: The case of tax auditing,” Journal of Econometrics, vol. 249, Art. no. 105847, 2025.

[10] S. G. Ayinaddis, "Artificial intelligence adoption dynamics and knowledge in SMEs and large firms: A systematic review and bibliometric analysis," Journal of Innovation & Knowledge, vol. 10, no. 3, Art. no. 100682, 2025.

[11] G. Qu and H. Jing, "Is new technology always good? Artificial intelligence and corporate tax avoidance: Evidence from China," International Review of Economics & Finance, vol. 98, Art. no. 103949, 2025.

[12] R. Belahouaoui and E. H. Attak, "Digital taxation, artificial intelligence and Tax Administration 3.0: improving tax compliance behavior-a systematic literature review using textometry (2016-2023)," Accounting Research Journal, vol. 37, no. 2, pp. 172-191, 2024.

[13] R. Belahouaoui and J. Alm, "Tax fraud detection using artificial intelligence-based technologies: trends and implications," Journal of Risk and Financial Management, vol. 18, no. 9, Art. no. 502, 2025.

[14] L. Oldemeyer, A. Jede, and F. Teuteberg, "Investigation of artificial intelligence in SMEs: a systematic review of the state of the art and the main implementation challenges," Management Review Quarterly, vol. 75, no. 2, pp. 1185-1227, 2025.

[15] M. Wang, "AI technologies in modern taxation: applications, challenges, and strategic directions," International Journal of Finance and Investment, vol. 1, no. 1, pp. 42-46, 2024.