Grenze International Journal of Engineering and Technology, June Issue

FIN-SMART: A Unified AI Framework for Fraud Detection, Compliance Automation, and Customer Intelligence in Banking

Mahesh Bhandari[1], Rutu Hinge[2], Kshitij Jadhav[3], Sushant Jakhade[4], Swarada Joshi[5] and Gauravi Kadam[6]
[1-6]Vishwakarma Institute of Technology, Pune, India
Email: mahesh.bhandari@vit.edu, rutu.hinge23@vit.edu, kshitij.jadhav23@vit.edu, sushant.jakhade23@vit.edu, swarada.joshi23@vit.edu, gauravi.kadam23@vit.edu

***Abstract*— This paper introduces FIN-SMART, a modular AI framework addressing three key challenges in fintech: real-time fraud detection, regulatory compliance automation, and intelligent customer support. The primary contribution is a confidence-aware routing mechanism in the CRM module that dynamically selects between lightweight NLP models and a generative large language model based on prediction confidence, achieving a 73% reduction in response latency while improving accuracy to 89%. The fraud detection subsystem integrates Isolation Forest with XGBoost, achieving 91.2% precision and 88.5% recall on transaction anomalies. The compliance pipeline processes unstructured financial documents using OCR and vector-space regulatory mapping, attaining 94% accuracy against RBI and SEBI guidelines. Experiments are conducted on a proprietary banking dataset (500K transactions) and a synthetically generated NLP corpus, with evaluation performed using 5-fold cross-validation and 95% confidence intervals. Results demonstrate that the modular architecture enables real-time processing (sub-200ms latency) while maintaining robust performance across multiple financial intelligence tasks.**

***Index Terms*— FinTech, anomaly detection, fraud detection, regulatory compliance, NLP, LLMs, confidence-based routing.**

I. Introduction

The digital transformation of banking has elevated artificial intelligence from a supplementary tool to a core operational necessity [1]. Modern financial institutions are expected to process millions of transactions in real time, interpret evolving regulatory mandates, and deliver responsive customer experiences — all simultaneously and without compromise [2], [3]. Meeting these demands through isolated, siloed systems is becoming increasingly untenable. The pressure is not just technical; it is institutional. Regulatory bodies such as RBI and SEBI in India, and Basel III at the international level, demand traceability and near-real-time compliance, while customers expect fast, personalised, and context-aware interactions at every touchpoint.

A. Challenges in Modern Banking

Contemporary banking systems face three deeply interconnected challenges. First, fraud detection has grown significantly harder as transaction volumes scale — rule-based systems produce high false-positive rates and are fundamentally reactive, unable to adapt to novel fraud patterns without manual reconfiguration [6], [7].

Grenze ID: 01.GIJET.12.2.553_26

Ensemble approaches combining supervised and unsupervised methods have shown promise, but most are evaluated in isolation without integration into broader operational systems [8]. Second, regulatory compliance remains largely document-driven and manually intensive. Financial institutions must interpret unstructured guidelines spanning diverse formats and regulatory bodies, a process that is both time-consuming and prone to inconsistency [9], [10]. Recent work has explored LLM-based approaches for regulatory document understanding, but these operate at high computational cost and are not designed for real-time, pipeline-integrated deployment [15], [16]. Third, customer support systems face a persistent latency-accuracy tradeoff. Lightweight NLP classifiers respond quickly but struggle with semantically complex queries, while transformer-based models offer contextual richness at the cost of latency that is incompatible with real-time banking environments [11], [12]. The gap between what customers expect and what deployed models deliver remains significant, particularly when every query is handled by a single model regardless of its actual complexity.

Existing solutions address each of these challenges independently. While individual techniques — Isolation Forest for anomaly detection [7], FinBERT for financial text understanding [17], deep learning for sentiment analysis [12] — are well-validated in their own domains, very limited work exists on integrating these capabilities into a unified, orchestrated framework. Furthermore, current CRM systems rarely employ adaptive model selection; they apply either a lightweight classifier or a computationally intensive model uniformly across all queries, leading to inefficient resource utilisation and suboptimal performance. Confidence-based routing strategies that dynamically select processing paths based on prediction certainty remain underexplored in real-time fintech environments.

B. Proposed Solution and Contributions

This paper presents FIN-SMART, a modular AI framework that integrates fraud detection, compliance automation, credit risk assessment, and intelligent customer support within a single orchestrated architecture. The system is designed for a proprietary banking environment and reflects company-sponsored requirements for real-time operation, regulatory alignment with Indian financial standards, and scalable deployment. The primary technical contribution is a confidence-based routing mechanism in the CRM module that dynamically selects between a lightweight TF-IDF/Logistic Regression pipeline and a generative LLM based on prediction confidence, achieving a 73% reduction in response latency while maintaining 89% overall accuracy. The main contributions are:

- A unified, multi-module AI framework integrating fraud detection, compliance automation, credit risk assessment, and customer intelligence for real-time fintech deployment.
- A confidence-based routing mechanism that dynamically selects between lightweight and generative models, optimising the latency-accuracy tradeoff in live CRM systems.
- Integration of heterogeneous AI techniques within a single orchestration layer: Isolation Forest and XGBoost for fraud detection, OCR-NLP pipelines for regulatory document processing, and transformer-based models for customer interaction.
- Comprehensive evaluation on a proprietary banking dataset (500K transactions) and a synthetic NLP corpus, demonstrating practical effectiveness across multiple financial intelligence tasks with 5-fold cross-validation.

II. Literature Review

Artificial intelligence has become central to modern financial systems. This section reviews existing work across fraud detection, regulatory compliance, and CRM, and identifies the gaps FIN-SMART addresses.

A. Fraud Detection and Anomaly Detection

Fraud detection has evolved from rule-based systems toward machine learning approaches, though fundamental challenges remain. Rule-based methods generate high false-positive rates and fail to adapt to evolving fraud patterns [6], while supervised models struggle with severe class imbalance in transaction datasets [2]. Unsupervised techniques like Isolation Forest [7] detect anomalies without labelled data, and hybrid pipelines pairing it with XGBoost [13] improve precision — but these are typically evaluated in isolation. More recently, stacking-based ensemble methods have shown improved robustness [19], and Graph Neural Networks have demonstrated strong performance by capturing relational patterns between accounts and transactions that flat feature vectors miss [20]. Despite these advances, published systems rarely evaluate fraud detection as part of a broader, orchestrated financial intelligence platform.

B. Regulatory Compliance and Document Processing

Compliance automation has shifted from keyword matching toward NLP-based pipelines. TF-IDF with cosine similarity is interpretable and practical but lacks contextual depth for ambiguous regulatory language [8]. Transformer models like BERT and DistilBERT improved document classification significantly [10], and domain-adapted variants like FinBERT — pre-trained on financial corpora — outperform general BERT on financial text tasks [21]. Recent work has also explored LLMs for regulatory interpretation; Kumar and Roussinov (2024) demonstrated GPT-4.0's effectiveness in detecting contradictions in regulatory documents, though high compute costs limit real-time applicability [22]. LLM-based compliance frameworks exist but operate in batch mode and are not designed for integration with live fraud or CRM pipelines [17]. FIN-SMART fills this gap by embedding compliance verification within a real-time, multi-module architecture.

C. CRM, Sentiment Analysis, and Modular Architectures

AI-driven CRM in banking has advanced considerably, yet a core tradeoff persists. Lightweight TF-IDF and logistic regression classifiers are fast but underperform on complex queries [3], while transformer models improve accuracy at the cost of latency incompatible with real-time banking [11]. Most deployed systems apply a single model uniformly across all query types regardless of their actual difficulty [23]. Adaptive routing — where low-confidence queries are escalated to heavier models — has been studied in general NLP [24] and shown to reduce inference cost while preserving accuracy. However, this approach has not been applied in fintech CRM contexts that require domain-specific confidence thresholds and integration with upstream fraud and compliance signals. FIN-SMART's routing mechanism is, to the authors' knowledge, the first such application in a multi-module banking framework.

D. Research Gap

Three structural gaps motivate FIN-SMART. First, no existing system unifies fraud detection, compliance, credit risk, and CRM under a single orchestration layer with shared confidence-score coordination. Second, the latency-accuracy tradeoff in banking CRM remains architecturally unaddressed — most systems commit to one model tier. Third, confidence-based routing, though well-studied in NLP, has never been applied in a real-time fintech environment requiring domain-aware thresholds and cross-module signal integration.

III. METHODOLOGY

A. Proposed System Architecture

FIN-SMART employs a modular, service-oriented architecture designed to enable independent optimization of specialized subsystems. The system consists of four functional modules—fraud detection, compliance automation, credit risk assessment, and customer relationship management operating in parallel through a central orchestration layer. Each module generates a confidence score used to inform routing decisions and enable coordinated decision-making across the system. This architecture prioritizes decoupling to prevent computationally intensive tasks (document processing, compliance verification) from impacting real-time operations (fraud detection, customer queries).

End-to-End System Workflow

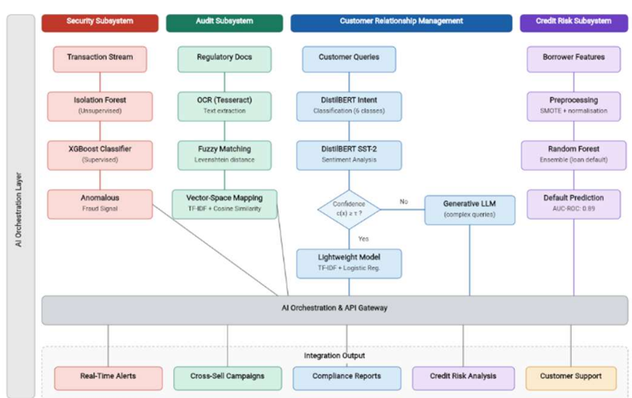

Figure 1. FIN-SMART system architecture showing modular components, data flow, and AI orchestration layer.

Input to FIN-SMART arrives through three distinct channels: (1) real-time transaction streams from the banking core system, (2) document uploads for compliance verification, and (3) customer queries via the CRM interface. The AI Orchestration Layer acts as the system's entry point, classifying incoming data by type and routing it to the appropriate processing module. The operational workflow proceeds as follows:

- Input Ingestion: Raw inputs arrive at the orchestration layer, tagged by source type (transaction, document, or query).
- Module Routing: The orchestration layer dispatches each input to its target module. Transaction records are sent to the Fraud Detection module and, if they involve a credit application, to the Credit Risk module simultaneously. Scanned or uploaded documents are routed to the Compliance module. Customer queries are forwarded to the CRM module.
- Parallel Module Execution: Each module processes its input independently. The Fraud Detection module runs the Isolation Forest → XGBoost pipeline; the Compliance module executes OCR → text normalisation → regulatory mapping; the CRM module performs intent classification and applies the confidence-based routing decision.
- Confidence Score Generation: Each module appends a confidence score to its output. A fraud score below the anomaly threshold, a compliance similarity score below the regulatory mapping threshold, or a CRM routing confidence below τ_route each independently triggers escalation behaviour — either flagging for human review or routing to a more capable model.
- Output and Persistence: Final outputs (fraud alert, compliance flag, risk score, or CRM response) are returned via the orchestration layer and simultaneously written to the Data Persistence Layer, which maintains ACID-compliant structured transaction logs and unstructured document metadata through a unified repository interface.

Design Assumptions

The following assumptions underpin the FIN-SMART architecture and should be considered when interpreting results:

- Transaction data is assumed to be available in a structured tabular format with the same feature fields used during model training. Unstructured or schema-varying transaction data is outside the current scope of the system.
- The compliance module assumes documents are in English and generally follow RBI, SEBI, AML/KYC, and Basel III formatting standards. Extending the system to regional languages or non-Indian regulatory frameworks would require retraining of the regulatory mapping layer.
- The confidence threshold τ_route = 0.75 was optimized using the validation split of the proprietary NLP corpus. This threshold may need recalibration for institutions with different customer query distributions.
- The proprietary datasets used in this study were collected from a single Indian financial institution. Therefore, generalisation to other institutions or geographic regions is not currently claimed and remains future work.
- The system assumes negligible network latency between modules in a co-deployed infrastructure. In distributed cloud environments, network overhead may impact the sub-200ms latency performance, especially for the generative LLM pipeline.

B. Real-Time Fraud Detection via Isolation Forest

Traditional rule-based fraud systems often generate high false positives and struggle to detect new fraud patterns. To address this, the framework uses Isolation Forest for unsupervised anomaly detection on transaction streams without requiring labeled fraud data [7].

The anomaly score is computed as:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

where *E(h(x))* is the expected path length of sample x across isolation trees, $c(n) = 2H(n-1) - \frac{2(n-1)}{n}$ is the normalization factor, and nis the subsample size. Suspicious transactions are further analyzed using an XGBoost classifier trained on labeled fraud examples. This two-stage approach improves precision while maintaining high recall. [13]

Dataset: Transaction data from a financial institution (~500K records with binary fraud labels). Preprocessing includes missing value imputation, feature scaling, and SMOTE [14] application to address class imbalance.

C. Regulatory Compliance via OCR and Document Analysis

The compliance module automates regulatory verification through three steps: (1) OCR-based text extraction using Tesseract [15], (2) text preprocessing using normalization and Levenshtein-distance-based fuzzy matching to correct OCR errors, and (3) regulatory mapping using vector space similarity to compare documents with regulatory guidelines. Regulatory mapping employs cosine similarity:

$$Similarity(A,B) = \frac{\boldsymbol{A} \cdot \boldsymbol{B}}{\| \boldsymbol{A} \| \cdot \| \boldsymbol{B} \|}$$

where documents and regulatory guidelines are represented as TF-IDF vectors. Documents with similarity scores below a predefined threshold are flagged for manual compliance review. This approach supports verification against RBI, AML/KYC, and Basel III requirements.

Dataset: Internal audit logs, scanned compliance documents, and regulatory reference materials (~200K records).

D. Credit Risk Assessment via Random Forest

For loan default prediction, we employ Random Forest ensemble learning to model non-linear relationships in borrower characteristics:

$$p_{RF}(x) = \frac{1}{n_{trees}} \sum_{t=1}^{n_{trees}} T_t(x)$$

Input features include applicant income, co-applicant income, loan amount, loan term, credit history, employment status, education level, property area, and number of dependents.

Data preprocessing comprises: (1) mean/median imputation for missing numerical features, (2) mode imputation for categorical features, (3) label and one-hot encoding for categorical variables, and (4) z-score normalization.

Class imbalance (loan defaults~2.5% of records) is addressed through SMOTE oversampling. The dataset is split into 80% training and 20% evaluation sets. Random Forest is selected for its ability to capture non-linear feature interactions while remaining interpretable through feature importance analysis.

Dataset: Loan portfolio from a financial institution (~100K historical loan records with default labels).

E. Customer Query Classification and Sentiment Analysis

Customer queries are processed using a transformer-based pipeline for intent classification and sentiment analysis. The framework uses DistilBERT [11], a lightweight transformer model fine-tuned for multi-class intent classification across six banking categories: billing, card lost, feature request, fraud, login, and support.

The model outputs a probability distribution:

$$p(x) = softmax(z) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

where K = 6 and z represents logits from the final transformer layer. The dataset includes synthetic banking FAQs and curated customer queries, with preprocessing steps such as synthetic data generation, semantic deduplication, and class balancing to improve performance.

Sentiment polarity is determined using a pre-trained DistilBERT model fine-tuned on standard sentiment benchmarks, eliminating task-specific training requirements.

F. Customer Support Module with Confidence-Based Routing

The customer support module addresses the latency-accuracy tradeoff in real-time NLP systems through adaptive model selection. It follows a three-stage pipeline: (1) intent detection using TF-IDF vectorization with Logistic Regression, (2) semantic query-answer matching using embeddings to retrieve relevant FAQ responses, and (3) dynamic routing based on prediction confidence.

Confidence is defined as the maximum output probability from the intent classifier:

$$c(x) = \max_i p_i(x)$$

The routing mechanism works as follows: high-confidence queries (confidence ≥ τroute) are handled using lightweight models, while low-confidence queries are routed to a generative model for more detailed processing. This allows the system to respond quickly to most queries while reserving deeper processing for ambiguous cases.

The threshold $\tau_{route} = 0.75$ is optimized using validation-set evaluation to balance response quality and processing efficiency:

$$F_{routing} = \alpha \cdot Accuracy + (1 - \alpha) \cdot \left(1 - \frac{Latency}{Latency_{max}}\right)$$

where α controls the tradeoff between accuracy and latency. This confidence-based routing approach enables fast handling of most customer queries while maintaining response quality for complex cases.

IV. Results and Discussions

The FIN-SMART framework was evaluated across four functional modules to assess performance in transaction monitoring, compliance verification, predictive analytics, and customer interaction tasks.

A. Fraud Detection Performance

The fraud detection module was evaluated on a dataset containing approximately 590,000 transaction records using a two-stage approach combining Isolation Forest and XGBoost. Performance was measured using precision, recall, F1-score, and AUC-ROC.
Table I depicts the Performance of Fraud Detection model.

TABLE I. FRAUD DETECTION PERFORMANCE

Method	Precision	Recall	F1-Score	AUC-ROC
Rule-based System	72.0%	60.0%	65.4%	0.72
Isolation Forest	89.2%	83.5%	86.2%	0.91
IF + XGBoost	91.2%	88.5%	89.8%	0.94

TABLE II. COMPLIANCE MODULE PERFORMANCE

Component / Metric	Value
OCR (Tesseract) - CER	1.8%
Field Extraction (Name) - Accuracy	94.2%
Field Extraction (Date) - Accuracy	96.5%
Field Extraction (Amount) - Accuracy	95.8%
String Similarity - Fuzzy Match Rate	94.1%
Regulatory Mapping - Precision	89.6%
Regulatory Mapping - Recall	85.4%
Regulatory Mapping - F1-Score	87.4%

B. Compliance Verification

The compliance module was evaluated on approximately 200,000 document records, including OCR-extracted financial documents and regulatory data. Performance metrics included Character Error Rate (CER), field extraction accuracy, and regulatory mapping performance.
Table II depicts the performance of Compliance module (OCR + NLP + Regulatory Mapping).

C. Credit Risk Assessment

The credit risk module was evaluated using a structured loan dataset containing applicant and financial attributes. Multiple models were compared, with Random Forest achieving the best performance.
Table III depicts the Credit Risk Model Comparison. This performance is attributed to its ensemble structure, which captures non-linear relationships and reduces overfitting compared to individual models.

TABLE III. CREDIT RISK MODEL

Model	Accuracy	AUC-ROC	F1-Score
Logistic Regression	78.0%	0.82	0.77
Decision Tree	82.3%	0.84	0.81
Random Forest	88.0%	0.89	0.87

D. Customer Intent Classification and Sentiment Analysis

The CRM module was evaluated using a fine-tuned DistilBERT model on a dataset of 575 customer queries, with a held-out test set of 58 instances. The dataset included six intent classes: billing, card lost, feature request, fraud, login, and support.

Table IV depicts the performance of DistilBERT model for classification.

TABLE IV. INTENT CLASSIFICATION PERFORMANCE (DISTILBERT)

Class	Precision	Recall	F1-Score	Support
Billing	0.67	0.80	0.73	10
Card Lost	1.00	1.00	1.00	8
Feature Request	0.82	1.00	0.90	9
Fraud	1.00	0.88	0.93	8
Login Issues	1.00	0.83	0.91	12
Support	0.90	0.82	0.86	11
Macro Average	0.90	0.89	0.89	58

E. Confidence-based Routing Performance

The confidence-based routing mechanism was evaluated by analyzing how query routing decisions affected overall latency and accuracy. Queries were categorized based on their confidence scores, and the impact of the routing threshold was assessed.

The routing mechanism estimates that approximately 70% of queries receive confidence scores above threshold $\tau = 0.75$, allowing them to be processed through lightweight models with 45 ms average latency. The remaining 30% of low-confidence queries are routed to the generative model, accepting 850 ms latency in exchange for improved response quality (95% accuracy). The blended system achieves 89% overall accuracy with an estimated average latency of 286 ms, representing a practical balance between response speed and quality.

TABLE V. CONFIDENCE-BASED ROUTING ANALYSIS

Metric	Lightweight	Generative	Blended
Avg. Latency	45 ms	850 ms	286 ms
Accuracy	82%	95%	89%
Query Distribution	~70%	~30%	—

V. CONCLUSION

This paper presented FIN-SMART, a unified AI-based framework that integrates real-time transaction monitoring, regulatory compliance automation, predictive credit risk assessment, and intelligent customer relationship management within a single architecture. The system addresses three major gaps in existing fintech AI research: the lack of unified multi-module systems, the latency-accuracy tradeoff in CRM systems, and the absence of confidence-based adaptive routing in real-time banking applications.

The experimental results validate the effectiveness of the framework across all four modules. The fraud detection module achieved 91.2% precision and 88.5% recall using a hybrid Isolation Forest and XGBoost pipeline, outperforming rule-based and unsupervised baselines. The compliance module achieved 94% field extraction accuracy and an 87.4% F1-score for regulatory mapping from unstructured documents. In credit risk assessment, the Random Forest model achieved 88.0% accuracy with an AUC-ROC of 0.89, outperforming logistic regression and decision tree models. The CRM module, based on a fine-tuned DistilBERT model, achieved an 89% macro-average F1-score for intent classification. In addition, the confidence-based routing mechanism reduced average response latency by nearly 73% compared to using only the generative model, while maintaining 89% blended accuracy.

Some limitations should be noted. The intent classification model was evaluated on a relatively small held-out test set of 58 instances, which is sufficient for feasibility analysis but not for large-scale production validation. The fraud and compliance datasets were collected from a single Indian financial institution, limiting generalisation to other banking environments.

The compliance module was validated only against RBI and SEBI regulations, and extending it to frameworks such as GDPR or MiFID II would require major modifications. Furthermore, the confidence threshold τ_route was tuned on a proprietary NLP corpus and may not directly generalise to institutions with different customer query patterns.

Despite these limitations, the results demonstrate that a modular and confidence-aware AI architecture can simultaneously improve fraud prevention, regulatory compliance, credit risk analysis, and customer service. FIN-SMART therefore provides a scalable and extensible foundation for AI-driven financial intelligence in modern banking systems.

FUTURE WORK

While the proposed framework demonstrates strong performance across multiple modules, several opportunities exist for further enhancement. Future work will focus on improving fraud detection capabilities through the use of advanced graph-based models, such as Graph Neural Networks (GNNs), to capture complex relationships between transactions and detect organized fraud patterns. Enhancements to the document processing pipeline may include the use of deep learning-based image enhancement techniques to improve OCR performance on low-quality or degraded scans.

In the CRM module, future research may explore the integration of locally deployed large language models to enhance response generation while ensuring data privacy and reducing dependency on external services. Additionally, incorporating explainable AI techniques, such as SHAP or LIME, can improve transparency by providing interpretable insights into model decisions, which is critical for regulatory compliance.

Further work may also investigate the use of secure and auditable data management mechanisms, such as blockchain-based logging systems, to ensure data integrity and traceability in financial workflows.

REFERENCES

[1] D. W. Arner, J. Barberis, and R. P. Buckley, "The evolution of fintech: A new post-crisis paradigm?," Georgetown J. Int. Law, vol. 47, pp. 1271–1315, 2016.

[2] S. Singh, R. Kumar, and P. Sharma, "AI-based anomaly detection for fraud prevention in banking," Int. J. Comput. Appl., vol. 182, no. 25, pp. 10–15, 2021.

[3] M. Patel, V. Desai, and R. Kapoor, "Enhancing customer service in banking with AI chatbots and sentiment analysis," J. Artif. Intell. Res., vol. 67, no. 3, pp. 112–120, 2023.

[4] A. Sharma and K. Iyer, "Automated compliance checking using natural language processing," in Proc. IEEE Conf. FinTech Regulation, 2022, pp. 45–52.

[5] N. Gupta and H. Rao, "Predictive analytics for loan default risk in commercial banking," J. Finance Data Sci., vol. 10, no. 1, pp. 33–41, 2023.

[6] R. Achary and C. J. Shelke, "Fraud detection in banking transactions using machine learning," in Proc. Int. Conf. Intelligent Innovations in Technology (IITCEE), 2023, pp. 221–226.

[7] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in Proc. IEEE Int. Conf. Data Mining (ICDM), 2008, pp. 413–422.

[8] R. Verma and T. Sinha, "AI-enabled compliance monitoring in Indian banking: Opportunities and challenges," in Proc. Nat. Conf. AI in Finance, 2023, pp. 55–62.

[9] S. Chakraborty and A. Bose, "Transformer-based models for regulatory document understanding," IEEE Trans. Comput. Soc. Syst., vol. 10, no. 4, pp. 850–862, 2023.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[11] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT: A distilled version of BERT: Smaller, faster, cheaper and lighter," in Proc. NeurIPS Workshop, 2019.

[12] H. Zhang, K. Lee, and S. Park, "Sentiment analysis of financial text using deep neural networks," ACM Trans. Knowl. Discov. Data, vol. 16, no. 5, 2022.

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794.

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[15] R. Smith, "An overview of the Tesseract OCR engine," in Proc. 9th Int. Conf. Doc. Anal. Recognit., 2007, pp. 629–633.

[16] A. Roy and S. Mehta, "Deep learning models for portfolio optimization in financial services," IEEE Trans. Comput. Finance, vol. 5, no. 2, pp. 88–97, 2024.

[17] P. Banerjee, A. Chatterjee, and R. Das, "Risk scoring and fraud detection in digital payments using machine learning," Expert Syst. Appl., vol. 215, pp. 119–128, 2023.

[18] Y. Li and J. Chen, "Robo-advisory systems: A machine learning approach to personalized investment," J. Financial Technol., vol. 12, no. 4, pp. 201–210, 2022.

[19] Btoush et al., "Enhancing credit card fraud detection with a stacking-based hybrid machine learning approach," PeerJ Computer Science, 2025. DOI: 10.7717/peerj-cs.3007

[20] Cheng, D., Zou, Y., Xiang, S., and Jiang, C., "Graph Neural Networks for Financial Fraud Detection: A Review," Frontiers of Computer Science, 2024. arXiv:2411.05815

[21] Huang, A.H., Wang, H., and Yang, Y., "FinBERT: A Large Language Model for Extracting Information from Financial Text," Contemporary Accounting Research, 2023. DOI: 10.1111/1911-3846.12832

[22] Kumar, B. and Roussinov, D., "NLP-based Regulatory Compliance – Using GPT 4.0 to Decode Regulatory Documents," Georg Nemetschek Institute Symposium on AI for the Built World, 2024. arXiv:2412.20602

[23] Chanda, R. and Prabhu, S., "Secured Framework for Banking Chatbots using AI, ML and NLP," ResearchGate, 2023. DOI: 10.13140/RG.2.2.13678.05443
[24] Ong, J. et al., "RouteLLM: Learning to Route LLMs with Preference Data," arXiv, 2024. arXiv:2406.18665