Forecasting Anomaly score for Financial Fraud Detection using Convolution Neural Networks

PVVS Eswara Rao[1], Rambabu Pasumarthy[2], MVB Murali Krishna M[3], Satya Srinivas Maddipati[2], Suresh Kumar Samarla[4], M Chilaka Rao[4]

[1] Assistant Professor, Sasi Institute of Technology & Engeering, Tadepalligudem, India
[2] Associate Professor, Sasi Institute of Technology & Engeering, Tadepalligudem, India
[3]Assistant Professor, Aditya University, Surampalem, India
[4]Assistant Professor, SRKR Engineering College, Bhimavaram, India

Abstract:- Financial statement fraud has become a critical concern, causing substantial economic losses to a wide range of stakeholders, including investors, governments, and financial institutions. The growing complexity and digitization of global financial systems have opened new avenues for fraudulent activities, challenging the effectiveness of traditional detection methods. To address this issue, the study proposes a Financial Fraud Detection Model (FFDM) leveraging Graph Neural Networks (GNNs) and Convolutional Neural Networks (CNNs). The model integrates both network-based and feature-based analysis to uncover hidden patterns and suspicious behaviours within dynamic transaction networks. Emphasis is placed on predicting anomaly scores using CNNs to proactively identify potential fraud before it occurs. The methodology incorporates mobile money platforms, credit card fraud, and fraudulent phone call detection as case studies. Experimental evaluation using various machine learning models demonstrates that CNNs outperform traditional methods like Decision Trees and Logistic Regression, achieving a significantly lower Mean Squared Error (MSE) in probability score prediction. This research highlights the potential of advanced AI techniques in enhancing the accuracy and speed of financial fraud detection systems.

Keywords*: Financial Fraud Detection, Convolution Neural Networks, Mean Squared Error, Probability Score*

1. Introduction

Financial statement fraud (FSF) has led to losses exceeding $500 billion for stakeholders such as investors, creditors, pensioners, and employees over the past few decades. In recent years, the global expansion of financial systems across corporations, governments, and academic institutions has introduced multiple avenues for fraud, including breaches involving computers, networks, customers, and internal personnel. These elements remain crucial in evaluating the risks associated with financial systems. Traditional approaches like audits and statistical models have limitations in accurately identifying the core features indicative of fraud within these systems. Financial fraud, particularly money laundering, represents a major criminal offense where unlawfully obtained funds are funnelled into activities like terrorism or other illicit operations. These crimes often involve intricate trade and financial transaction networks, making it challenging to pinpoint fraudulent entities and uncover the patterns behind such actions. However, from these complex networks, one can construct a transaction network and extract entity-specific features. The transaction network illustrates interactions among entities, enabling anomaly detection techniques to identify those potentially involved in fraudulent activities. On the other hand, analysing entity features helps to expose specific characteristics linked to fraud. Together, network structures and entity features offer complementary insights that can significantly enhance the accuracy of fraud detection.

Financial fraud remains a persistent and escalating threat with significant implications for the financial sector. Machine learning has become an essential tool in detecting fraudulent activities within financial transactions. Despite this progress, fraud detection continues to be a complex task for two primary reasons: the rapid and continuous evolution of both fraudulent and legitimate transaction patterns, and the high-speed nature of modern online transactions, which demands swift and precise detection mechanisms.

1.1 Mobile money platforms

Mobile money platforms have become a transformative force in financial technology, especially in developing regions, by offering banking services to populations that previously lacked access. Despite their benefits, these digital systems have also become targets for increasingly complex fraudulent activities, posing major challenges for financial institutions globally. Detecting financial fraud is vital for ensuring secure financial operations and reducing the risk of illicit activities.

1.2 Fraudulent phone calls

Fraudulent phone calls have become a widespread issue, resulting in major financial damage and posing serious risks to both individuals and organizations. This project aims to create a dedicated platform for detecting and preventing such fraudulent calls. By utilizing cutting-edge technologies like machine learning and anomaly detection, the platform is designed to identify and intercept suspicious calls in real time. The rapid spread of information via digital platforms has transformed how data is accessed and utilized, but it has also contributed to a rise in financial statement fraud, threatening the integrity and efficiency of capital markets.

The Financial Fraud Detection Model (FFDM) presents a sophisticated approach to pinpointing fraudulent financial transactions using Graph Neural Networks (GNNs). With the rapid growth in the number of credit card users and companies conducting financial transactions via credit cards, there has been a significant rise in fraudulent activities. This issue is further complicated by the presence of noisy, imbalanced data and outliers.

The growing concern around financial fraud has led to heightened interest from industry experts, resulting in the ongoing advancement of detection techniques. Most existing approaches treat customer data as isolated points, overlooking the fact that fraudulent actions often share common traits and are frequently carried out by coordinated groups. Dynamic graph-based fraud detection focuses on identifying anomalous entities that significantly diverge from typical benign behaviour within evolving network structures. However, current approaches often struggle to adapt effectively across diverse financial fraud scenarios, limiting their ability to ensure financial system security. Real-time financial transactions exhibit strong temporal correlations, especially among fraudulent activities, which pose additional challenges. Conventional dynamic graph models typically fail to balance the representation of temporal dynamics and spatial structure, and are inadequate when handling heterogeneous node types within the network.

1.3 Digital Platforms:

The widespread use of digital platforms has dramatically transformed how data is accessed and consumed, inadvertently creating an environment conducive to financial statement fraud. Such fraud poses a serious threat to the effective functioning of capital markets. The rapid growth of Internet finance has led to a surge in financial fraud incidents, posing significant threats to the stability and security of the financial sector. The increasing reliance on electronic payment systems has elevated the critical need for effective financial fraud detection, particularly in the realm of credit card fraud. This task is complicated by challenges such as class imbalance, overfitting, and the difficulty of accurately capturing the intricate and evolving nature of fraudulent activities.

2. Literature Survey

The study[2] examines several forms of fraud, including payment fraud and account takeovers, stressing the growing concern as online payment fraud is expected to surpass $343 billion by 2027. Authors investigated a range of detection techniques such as Machine Learning, Behavioural Analysis, Anomaly Detection, and Data Analytics. Using a synthetic dataset based on mobile money transactions, they analysed the performance of three machine learning algorithms: Logistic Regression, Decision Tree, and Multi-Layer Perceptron (MLP). Logistic

Regression faced challenges due to data imbalance, whereas the Decision Tree algorithm provided more consistent results in terms of precision and recall. Although the MLP model delivered high precision, it had lower recall, leading to undetected fraud cases. These outcomes underline the importance of optimizing models for better fraud detection accuracy. The research highlights key security risks in finance and suggests adaptive solutions to safeguard both businesses and consumers in today's digital economy. The study[3] centres on identity theft within financial systems and introduces a comprehensive framework that combines both qualitative and quantitative techniques for detecting fraudulent activities. The framework integrates the Analytic Hierarchy Process (AHP) with Rough Set (RS) theory to evaluate fraud cases, identify key features, detect anomalies, and generate alerts. While designed to target identity theft, the framework is also adaptable for detecting various forms of fraud within financial systems.

The model presented in [1] introduces an innovative GNN-powered solution for financial fraud detection. It merges graph-based learning with deep neural networks to effectively uncover the intricate connections and behaviours within financial transaction data. Through a multi-layered design, the GNN model excels at recognizing unusual patterns associated with fraud and can adjust to new deceptive tactics. A significant advantage of this model is its incorporation of both node and edge features, which enriches the representation of transaction networks.

The research[4] presents a robust fraud detection framework tailored for mobile money transactions, leveraging advanced machine learning methods. A synthetic dataset modeled on actual financial transaction logs—generated using PaySim, a simulation tool based on data from a mobile money service in an African nation—was used for experimentation. Three widely-used machine learning models were applied and assessed: XGBoost, Logistic Regression, and Random Forest. Among these, the Random Forest classifier demonstrated superior performance, achieving an outstanding Area Under the Precision-Recall Curve (AUPRC) of 0.9998. Additionally, the implementation of feature engineering, including the use of error balance metrics, significantly improved the accuracy of fraud detection. The findings underscore the importance of developing sophisticated detection techniques, which hold substantial value in enhancing the security and trustworthiness of digital financial transactions. Continued innovation in this area can greatly contribute to building safer and more dependable financial ecosystems.

The project [5] involves in-depth analysis of current fraud detection techniques, the development of an intuitive interface for users to report and monitor suspicious activities, and the incorporation of powerful algorithms to improve detection precision. The primary goal is to minimize the occurrence of financial fraud, safeguard users from potential scams, and foster a more secure communication landscape. The insights and practical tools developed through this initiative are intended to support ongoing efforts to combat fraud in the telecom industry. The study [6] introduces an advanced representation learning approach aimed at identifying such fraud by analysing the evolution of a company's Management Discussion and Analysis (MD&A) sections over time. Instead of relying on conventional word frequency methods, their approach aligns paragraphs across successive reports by measuring their similarity at the representation level. Based on this alignment, paragraphs are classified into three categories: added, removed, and retained. Authors then generate multivariate change trajectory features using fraud-related linguistic categories, including sentiment and expressions of uncertainty. These features are used to train a fraud detection model that captures nuanced textual changes over time. Using 24 years of financial filings from 1995 to 2019, our experiments demonstrate that this representation learning strategy consistently outperforms traditional models across seven machine learning frameworks. This approach introduces a novel direction in feature engineering for detecting financial statement fraud.

Graph Neural Networks (GNNs) have demonstrated promising outcomes in this domain. However, challenges such as limited data exploitation and class imbalance within heterogeneous financial transaction graphs still persist. To address these issues, the study [6] introduces Meta path Graph Neural Networks (Metapath-GNN)—a specialized GNN framework designed for uncovering financial fraud within complex transaction environments and hidden behavioural patterns. This model begins by constructing subgraphs based on predefined meta path structures using a dedicated generation module. An attention mechanism is then applied to refine node selection, helping to mitigate the effects of class imbalance. Following this, an aggregation module integrates information from both subgraphs and the overall graph, resulting in richer and more informative node representations. This allows the model to leverage crucial information effectively, enhancing its fraud detection capabilities. The performance of Metapath-GNN is thoroughly evaluated using public datasets such as YelpChi, Amazon, and Elliptic. Notably, for the real-world financial transaction dataset Elliptic, the model employs a semi-supervised learning technique to reduce the need for labelled data by incorporating unlabelled data in training. Experimental comparisons with state-of-the-art methods show that Metapath-GNN achieves superior results. This technique improved F1-macro, AUC, and GMean On YelpChi, Amazon, Elliptic and T-Finance datasets. Despite advancements in fraud detection, several challenges persist, including limited model interpretability, slow adaptation to emerging fraud tactics, and insufficient data mining. To tackle these issues, authors introduced[7] a novel graph-based learning model named MetaFraud-GNN (Metagraph Fraud Detection Graph Neural Network). This framework integrates metagraph search and neural architecture search (NAS) to automatically refine the network design for enhanced fraud detection in financial transaction systems.MetaFraud-GNN utilizes metagraph search algorithms to uncover intricate subgraph patterns that frequently indicate fraudulent behavior. By identifying and leveraging these key structural motifs, the model can extract deeper and more nuanced insights from transaction data, improving its ability to detect hidden fraudulent activity. Furthermore, a metagraph decoding mechanism is employed to dynamically optimize the graph neural network’s architecture, ensuring the model remains effective against constantly evolving fraud strategies. This technique was evaluated the performance of MetaFraud-GNN on three widely used real-world datasets,YelpChi, Amazon, and Elliptic and find that it consistently outperforms existing state-of-the-art models. The improvements are evident across multiple evaluation criteria, including F1-macro, Area Under the Curve (AUC), and Geometric Mean (GMean), demonstrating the model’s superior accuracy and adaptability in fraud detection tasks.

Effective financial fraud detection is essential for safeguarding assets and ensuring the stability of the financial system. Traditional detection approaches often lack adaptability, while many machine learning techniques, despite their accuracy, are typically complex and hard to interpret. To address this, Wenjuan Li et. al, introduced the XGB-GP framework, which integrates Extreme Gradient Boosting (XGB) with Genetic Programming (GP) to develop interpretable and efficient fraud detection models. This hybrid approach not only enhances model transparency but also boosts detection performance. Their analysis identifies the financial ratio “Total Liabilities to Operating Costs” as a key indicator of fraudulent activity. Experiments conducted using data from the CSMAR database, which includes Chinese publicly listed firms, show that XGB-GP significantly surpasses both conventional and machine learning models in identifying financial fraud [8]. Corporate governance metrics are critical in enhancing the detection of financial fraud. While the Extreme Gradient Boosting (XGBoost) model demonstrates greater reliability than the Multilayer Perceptron Neural Network (MLP NN) in identifying fraudulent activities, it faces challenges in parameter tuning. To overcome this, the Ant Colony Optimization algorithm is employed to fine-tune the model, significantly improving its accuracy. Utilizing data from 1,660 publicly listed Chinese companies spanning 2015 to 2021, the incorporation of corporate governance features notably boosted the predictive performance of the XGBoost model. Both model optimization and empirical findings reveal that fraud detection is more accurate in the initial phases of a fraud cycle than in later stages. Furthermore, shorter fraud learning cycles yield better detection results compared to longer ones, with a two-year cycle emerging as the most effective. The study [9] also explores the underlying mechanisms through which corporate governance factors contribute to improved financial fraud detection.

Most existing approaches rely solely on either network or feature-based data, missing the benefits of a combined perspective. To address this limitation[10], this paper introduces a novel detection framework called CoDetect, which integrates both network and feature data to identify financial fraud. Moreover, CoDetect not only detects fraudulent entities but also uncovers the distinctive feature patterns associated with fraudulent behavior. Experimental results on both synthetic and real-world datasets validate the effectiveness and robustness of our approach, particularly in tackling money laundering.

As wireless communication systems become essential for handling large-scale data transfers and safeguarding against interference, the escalating threat of financial fraud has emerged as a pressing issue. To address this challenge, A. A. Almazroi et. al introduced [11] a novel AI-based framework cantered around the ResNeXt-integrated Gated Recurrent Unit (GRU) architecture, referred to as RXT, specifically designed for real-time analysis of financial transactions. Driven by the urgent need to protect both financial institutions and their customers from fraudulent activity, their approach employs a structured AI pipeline. It begins with data ingestion and preprocessing, including the use of SMOTE to correct class imbalances. For feature extraction, authors deployed a hybrid ensemble technique that combines autoencoders with ResNet, termed EARN, to capture meaningful patterns in the data, while feature engineering further sharpens the model's capacity to differentiate between normal and suspicious transactions. At the heart of our classification process lies the RXT model, which is fine-tuned through the Jaya optimization algorithm (yielding RXT-J) for enhanced accuracy. Rigorous testing on three real-world financial datasets demonstrates that their method surpasses existing techniques, achieving 10% to 18% higher performance across various evaluation metrics, all while maintaining computational efficiency. This cutting-edge AI solution marks a significant step forward in combating financial fraud and bolstering the security, availability, and reliability of financial systems, particularly in the face of interference and cyber threats targeting wireless communication infrastructure. The research [12] investigates the effectiveness of various machine learning models in detecting financial statement fraud and evaluates how specific financial indicators influence model performance. Using data from fraud cases reported by the U.S. Securities and Exchange Commission (SEC) between 2016 and 2019 (publicly released from 2021 to 2023), the study utilizes fifteen selected financial indicators as features. Five classification algorithms—Decision Tree, Logistic Regression, Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost)—are employed for training and evaluation. To mitigate class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied. The findings suggest that XGBoost and SVM are the most effective at detecting fraudulent statements, although SVM may be prone to overfitting. Random Forest demonstrates consistent and reliable performance. Among the financial indicators, Interest-Bearing Debt to Total Invested Capital (IBD/TIC), Quick Ratio (QR), Accounts Payable Turnover Ratio (APTR), Goodwill Proportion (GP), and Goodwill (GW) emerge as influential variables, playing a key role in fraud detection across several models. A major contribution of this work lies in its focus on model interpretability through financial indicators, offering valuable insights for practical fraud detection. This study supports the advancement of more accurate and transparent fraud detection systems and provides meaningful guidance for auditors, financial professionals, and regulatory bodies by aligning theoretical methods with practical application. Incidents of financial fraud that severely impact investors are fairly frequent. To address this issue, a variety of intelligent detection approaches have been introduced [13] to assist financial institutions in making informed decisions. However, many of the existing methods face limitations such as low detection accuracy, slow processing times, and poor generalization across diverse datasets. To overcome these challenges, Yuxuan Tang et. al proposed [14] a Transformer-based distributed knowledge distillation framework for financial fraud detection. The approach begins by applying a multi-attention mechanism to assign importance weights to features, followed by feed-forward neural networks to extract meaningful high-level representations. These are then used by classification layers to identify fraudulent activities. Additionally, to tackle challenges such as inconsistent financial metrics across industries and imbalanced data, a distributed knowledge distillation algorithm is developed. This algorithm integrates detection knowledge from multiple teacher networks and transfers it to a student network that is capable of handling industry-specific financial data. Experimental evaluations reveal that this method significantly outperforms traditional techniques, achieving an F1 score of 92.87%, an accuracy of 98.98%, a precision of 81.48%, a recall of 95.45%, and an AUC of 96.73%.

The study [15] introduced an advanced representation learning technique aimed at identifying financial statement fraud by analysing temporal changes in the Management Discussion and Analysis (MD&A) sections of company reports. Unlike conventional methods that rely solely on word frequency, this approach begins by aligning paragraphs from consecutive reports using their semantic similarity. Based on these alignments, paragraphs are categorized as added, removed, or unchanged. Authors then generate multivariate trajectories of change using lexicons relevant to fraud detection, including those focused on sentiment and uncertainty. These trajectories serve as inputs for our fraud detection model. Comprehensive evaluations using financial disclosures from 1995 to 2019 demonstrate that our method substantially enhances fraud detection accuracy across seven machine learning models, consistently outperforming traditional frequency-based techniques. This work established a novel direction for feature engineering in the context of financial statement fraud detection.

Binbin Fang et.al introduced [16] GraphFA, an innovative graph-based fraud detection framework that incorporates camouflage detection. This method begins by representing customer data as a graph structure that reflects inter-customer relationships, framing the fraud detection task as a node classification problem. To tackle the challenge of hidden or misleading connections, referred to as camouflaged edges, that degrade the effectiveness of graph-based detectors on large-scale networks, GraphFA includes a neural sampler specifically designed to identify and handle these deceptive edges during subgraph extraction. As far as we are aware, GraphFA is the first solution in financial fraud detection to combine graph modelling with camouflage edge detection. Experimental evaluation on a real-world loan fraud dataset demonstrates that GraphFA surpasses both advanced models and widely adopted gradient boosting decision tree techniques. Notably, it achieves an AVC of 90.5% and a 4.7% improvement in Recall for individual fraud detection. On datasets involving group fraud, it records a 7.5% gain in F1-score and a 5.9% rise in Recall over the strongest baseline. The study [18] introduced a novel framework named Dynamic Heterogeneous Transaction Graph Embedding (DyHDGE). Built upon dynamic heterogeneous transaction graphs, this approach simultaneously captures both time-dependent and structural patterns while accommodating various data types. Specifically, Xinzhi Wang et. al, implemented two specialized modules: one to learn temporal relationships within transactions and another to extract spatial structural features across nodes. Furthermore, authors et.al, designed dual loss functions to refine node feature representations effectively. Experiments conducted on two synthetic financial fraud datasets demonstrate that DyHDGE achieves superior performance compared to existing state-of-the-art techniques, offering improved security and detection accuracy in financial transaction environments.

The study [17] presents a fraud detection approach utilizing the CatBoost machine learning algorithm. To enhance classification performance, Yeming Chen et. al, implemented feature engineering to extract and construct highly informative features, which are then used as input for CatBoost. A distinctive aspect of their work is the incorporation of memory compression techniques to accelerate the detection process. This method is tested on the IEEE-CIS Fraud Detection dataset from the Kaggle platform. Experimental results reveal that their CatBoost-based model achieves a peak accuracy of 0.983, showcasing its effectiveness.

The research [19] investigates the impact of internal and external perceptions, along with the disparities between them, on improving the accuracy of financial fraud detection. This study develops and evaluates management and media perception metrics using advanced analytical techniques. Drawing on data from the Chinese market spanning 2017 to 2021, the findings indicate that incorporating perspectives from multiple stakeholders significantly enhances detection performance. Additional cross-sectional analyses and real-world applications further validate these insights. This study also explores the predictive relevance and broader implications of perception-based indicators. Notably, the findings suggest that media outlets are capable of identifying potentially fraudulent corporate behaviour and may serve a regulatory function in monitoring firms.

The study [20] focuses on leveraging artificial intelligence (AI) to enhance data mining approaches for developing an effective system for detecting and preventing financial fraud. Initially, this research explores the evolution of Internet finance and the growing complexity of fraud cases, highlighting the inadequacies of traditional prevention strategies in the era of big data. It then delves into the promising role of AI in the financial industry, emphasizing its strengths in data analysis and fraud detection. This study thoroughly examines the data mining methods employed, including AI-driven and machine learning-based models, as well as techniques for data preprocessing and feature engineering within large-scale data environments. Furthermore, it outlines the system's architecture and the core technologies implemented, such as model training, performance optimization, real-time surveillance, and early warning mechanisms. Finally, this study presents a detailed assessment of experimental outcomes, confirming the system’s high effectiveness and practical viability in combating financial fraud. The findings demonstrate that the model performs well on real-world financial datasets, offering strong technical support and risk mitigation capabilities for the financial industry. By utilizing the capabilities of generative adversarial networks (GANs), the model [21] can synthesize realistic fraud-like data, which contributes to more robust and effective training of detection algorithms. Experimental evaluations on multiple credit card transaction datasets demonstrate that the SAGAN-based method achieves notable improvements in detection accuracy and recall. This model excels at capturing the nuanced and diverse characteristics of fraudulent activity, significantly boosting both the precision and overall performance of fraud detection systems.

In the study [22], Rohan Y.G. et.al, tackled these challenges by applying three data imbalance handling methods alongside six different classification algorithms. Additionally, six types of neural network models were employed. The research utilized real-world data from Ayushman Bharat (PM-JAY), India’s flagship public health insurance scheme and one of the largest in the world. A total of 26 model configurations were evaluated using performance metrics such as accuracy, sensitivity, specificity, and F1-score. Among these, a neural network model trained on an under sampled dataset showed superior performance compared to the other models tested.

The study [23] introduced an artificial intelligence-based approach for credit card fraud detection. The system utilizes logistic regression to construct a classification model aimed at identifying and preventing fraudulent transactions. To enhance the model's accuracy and reliability, a data pre-processing stage is incorporated. This stage employs two innovative techniques for data cleaning: a mean-based method and a clustering-based method. When compared with established models such as the support vector machine (SVM) classifier and the voting classifier, the proposed logistic regression model demonstrated superior performance in terms of detection accuracy, sensitivity, and error reduction.

3. Methodology

In financial fraud detection, one common approach is to assign an anomaly score to transactions, this score indicates how unusual or suspicious a transaction appears when compared to typical, legitimate behaviour. The higher the score, the more likely the transaction is to be fraudulent.

Forecasting anomaly scores means predicting these scores in advance, based on patterns learned from historical financial data. Instead of simply reacting to already-occurred transactions, the system tries to anticipate suspicious activity before or as it happens.

To accomplish this, Convolutional Neural Networks (CNNs), typically used in image processing, are adapted for analysing structured financial data. Although CNNs are best known for detecting spatial patterns in images, they can also learn and extract meaningful features from tabular or time-series data. When financial data is organized like a matrix (for example, as sequences of transaction features over time), CNNs can identify complex patterns and relationships that indicate fraudulent behaviour.

In this context, the CNN learns from past labelled data (fraudulent and non-fraudulent transactions) and predicts anomaly scores for new transactions. These scores help financial institutions prioritize which transactions need further investigation or real-time blocking. Figure 3.1 presents proposed methodology.

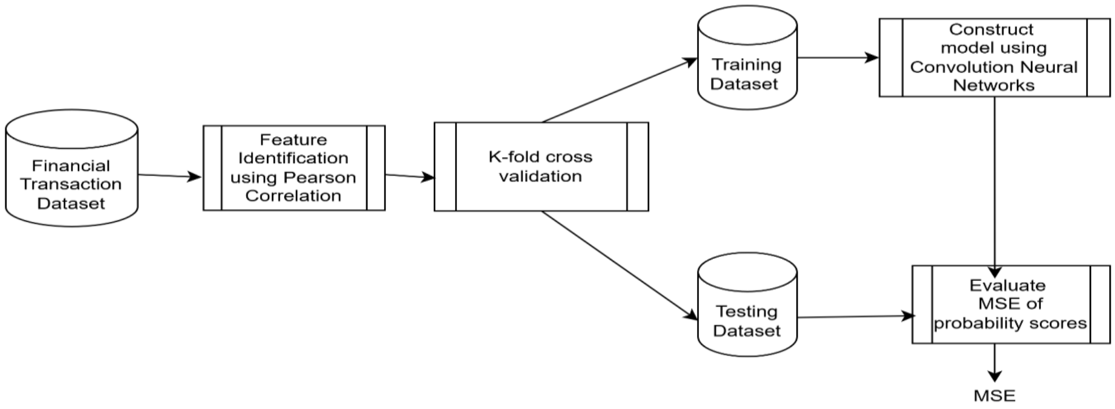
Financial Transaction Dataset
Feature Identification using Pearson Correlation
K-fold cross validation
Training Dataset
Construct model using Convolution Neural Networks
Testing Dataset
Evaluate MSE of probability scores
MSE

Figure 3.1: Architecture diagram of Financial Fraud detection

4. Results

The proposed model can be applied to a dataset to generate either classification results or scores for each individual entry. Typically, the testing dataset is used for scoring. Alternatively, a user can upload a CSV file for scoring purposes. Figure 4.1 shows the results of Pearson correlation on Financial Fraud detection. Q-Q residuals refer to a graphical method used to assess whether the residuals (errors) from a statistical model follow a particular theoretical distribution, most commonly, the normal distribution. Figure 4.2 presents Q-Q residuals. For binary classification models, the system can output a probability score indicating the likelihood of a specific class. In regression models, a numerical prediction is generated for each entry. For other types of models, a categorical class label is assigned to each data point. Users can manage these outputs through the settings related to Class and Probability.

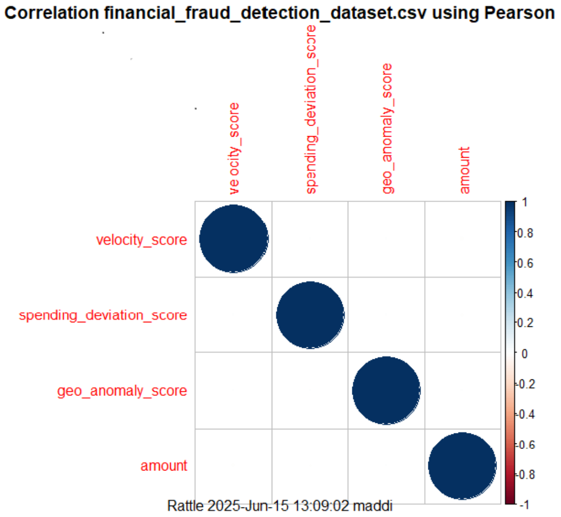

Figure 4.1 : Pearson Correlation

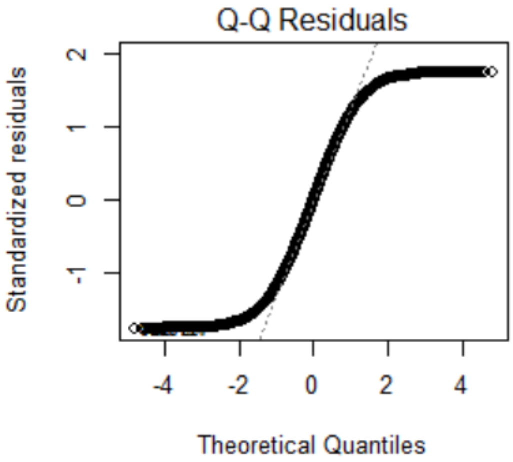

Fig 4.2: Q-Q residuals

The MSE measures the average squared difference between the predicted probability score (from the model), and the actual value. Table 4.1 presents Mean Squared Error (MSE) values of probability scores on Financial Fraud detection

Table 4.1: MSE values of Probability scores

Model	MSE of probability score
Decision Trees	1.21
Logistic Regression	1.005
Convolution Neural Networks	0.235

5. Concluusion

The reviewed studies collectively highlight the evolving landscape of financial fraud detection, where traditional approaches are increasingly being replaced or enhanced by advanced artificial intelligence and machine learning techniques. Research demonstrates that integrating internal and external perceptions, particularly from management and media, provides valuable insights that improve detection accuracy. At the same time, AI-driven frameworks, especially those employing data mining, neural networks, and generative adversarial networks, have shown significant promise in addressing the challenges of large-scale and complex fraud scenarios. Empirical evidence across diverse domains, including credit card transactions, healthcare insurance schemes, and capital markets, confirms the superiority of intelligent, perception-based, and data-driven models over conventional methods. The proposed methodology builds on these advancements by employing Convolutional Neural Networks (CNNs) to forecast anomaly scores, enabling proactive detection of suspicious transactions. Results show that CNN-based models considerably reduce error rates and outperform traditional algorithms, as reflected by the lowest MSE values compared to decision trees and logistic regression. These findings underscore the potential of deep learning architectures to capture intricate patterns in financial data and provide scalable, real-time fraud detection solutions.quis.

Refrences

[1] A. Kesharwani and P. Shukla, "FFDM − GNN:A Financial Fraud Detection Model using Graph Neural Network," 2024 International Conference on Computing, Sciences and Communications (ICCSC), Ghaziabad, India, 2024, pp. 1-6, doi: 10.1109/ICCSC62048.2024.10830438.

[2] M. Marripudugala, "AI-Powered Fraud Detection in the Financial Services Sector: A Machine Learning Approach," 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2024, pp. 795-799, doi: 10.1109/ICSSAS64001.2024.10760599.

[3] Qian Liu, Tong Li and Wei Xu, "A subjective and objective integrated method for fraud detection in financial systems," 2009 International Conference on Machine Learning and Cybernetics, Hebei, 2009, pp. 1339-1345, doi: 10.1109/ICMLC.2009.5212307.

[4] A. F. Sariat, I. J. Siddique, M. Hossain, M. M. Islam and T. Rahman, "AI Driven Fraud Detection in Financial Ecosystems: A Hybrid Machine Learning Framework," 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2025, pp. 1-8, doi: 10.1109/ECCE64574.2025.11013808.

[5] V. Tiwari and A. Pratap, "Fraud Call Detection using Pre-Data Feeding Method by Financial Institution," 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2025, pp. 990-993, doi: 10.1109/CICTN64563.2025.10932609.

[6] Junjie Qian, Guoxiang Tong, “Metapath-guided graph neural networks for financial fraud detection”, Computers and Electrical Engineering, Volume 126, 2025, 110428, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2025.110428.

[7] Guoxiang Tong, Junjie Qian, Jieyu Shen, “Adaptive metagraph neural network assisted by metagraph search for financial fraud detection”, Engineering Applications of Artificial Intelligence, Volume 153, 2025, 110807, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2025.110807.

[8] Wenjuan Li, Xinghua Liu, Junqi Su, Tianxiang Cui, “Advancing financial risk management: A transparent framework for effective fraud detection”, Finance Research Letters, Volume 75, 2025, 106865, ISSN 1544-6123, https://doi.org/10.1016/j.frl.2025.106865.

[9] Jing Li, “Corporate governance, fraud learning cycles, and financial fraud detection: Evidence from Chinese listed firms”, Research in International Business and Finance, Volume 76, 2025, 102832, ISSN 0275-5319, https://doi.org/10.1016/j.ribaf.2025.102832.

[10] D. Huang, D. Mu, L. Yang and X. Cai, "CoDetect: Financial Fraud Detection With Anomaly Feature Detection," in IEEE Access, vol. 6, pp. 19161-19174, 2018, doi: 10.1109/ACCESS.2018.2816564.

[11] A. A. Almazroi and N. Ayub, "Online Payment Fraud Detection Model Using Machine Learning Techniques," in IEEE Access, vol. 11, pp. 137188-137203, 2023, doi: 10.1109/ACCESS.2023.3339226.

[12] B. Li, J. Yen and S. Wang, "Uncovering Financial Statement Fraud: A Machine Learning Approach With Key Financial Indicators and Real-World Applications," in IEEE Access, vol. 12, pp. 194859-194870, 2024, doi: 10.1109/ACCESS.2024.3520249.

[13] Y. Tang and Z. Liu, "A Distributed Knowledge Distillation Framework for Financial Fraud Detection Based on Transformer," in IEEE Access, vol. 12, pp. 62899-62911, 2024, doi: 10.1109/ACCESS.2024.3387841.

[14] Tang, Yuxuan & Liu, Zhanjun. (2024). A Distributed Knowledge Distillation Framework for Financial Fraud Detection Based on Transformer. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3387841.

[15] Y. Yu, Z. Wu, Y. Han, Z. Li and W. Wei, "Unlocking Financial Statement Fraud Detection: Tracking Disclosure Changes via Representation Learning," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10887753.

[16] B. Fang, H. Chen, W. Wang and Y. Wang, "GraphFA: Graph Enhanced Fraud Detectors with Camouflage Detection for Financial Anti-Fraud," 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP), Xian, China, 2024, pp. 323-327, doi: 10.1109/ICSP62122.2024.10743929.

[17] Y. Chen and X. Han, "CatBoost for Fraud Detection in Financial Transactions," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2021, pp. 176-179, doi: 10.1109/ICCECE51280.2021.9342475.

[18] Xinzhi Wang, Jiayu Guo, Xiangfeng Luo, Hang Yu, “DyHDGE: Dynamic heterogeneous transaction graph embedding for safety-centric fraud detection in financial scenarios”, Journal of Safety Science and Resilience, 5(4), ISSN 2666-4496.

[19] Guowen Li, Shuai Wang, Yuyao Feng, “Making differences work: Financial fraud detection based on multi-subject perceptions”, Emerging Markets Review, Volume 60,2024, 101134, ISSN 1566-0141.

[20] Ziyue Wang, Qinyan Shen, Shuochen Bi, Chengqian Fu, ”AI Empowers Data Mining Models for Financial Fraud Detection and Prevention Systems”, Procedia Computer Science, Volume 243, 2024, ISSN 1877-0509.

[21] Chuanjun Zhao, Xuzhuang Sun, Meiling Wu, Lu Kang, “Advancing financial fraud detection: Self-attention generative adversarial networks for precise and effective identification”, Finance Research Letters, Volume 60, 2024, 104843, ISSN 1544-6123. .

[22] Rohan Yashraj Gupta, Satya Sai Mudigonda, Pallav Kumar Baruah "A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes" International Journal of Engineering Trends and Technology 69.3(2021):96-102.

[23] Hala Z Alenzi1, Nojood O Aljehane, “Fraud Detection in Credit Cards using Logistic Regression”, International Journal of Advanced Computer Science and Applications, 11(20),2020.