Iranian Journal of Operations Research
Vol. 16, No. 2, 2025, pp. 92-104

Intelligent detection of fraud in financial statements using deep learning and XGBoost

Mehdi Farrokhbakht Foumani [1,*] , Ali Akbar Akhavan [2]

Fraud is a phenomenon that involves deviations and manipulations in financial statements. These actions can lead to tax non-compliance and erode the trust stakeholders. Given and vast amount of financial data within organizations, leveraging artificial intelligence as a sophisticated tool can greatly enhance fraud detection in financial statements and bolster confidence in the face of evolving fraudulent tactics. This paper introduces an intelligent method for detecting fraud in financial statements. Initially, the Apriori algorithm is utilized to select pertinent features in the financial data. Subsequently, the performance of the proposed method is enhanced by augmenting the dataset using the GAN-CNN network. Finally, fraud detection is executed with the assistance of XGBoost. The proposed model is evaluated on a comprehensive financial dataset from Kaggle. The results demonstrate that the proposed method achieves superior performance with an accuracy of 96.21%, a precision of 96.02%, a recall of 95.41%, and an F1-score of 93.99%, significantly outperforming benchmark models such as 1D-CNN, LSTM, and Bi-LSTM. Furthermore, a sensitivity analysis reveals that the model maintains robust performance (accuracy >94%) across varying training data sizes and is not overly sensitive to the key hyperparameters of the GAN and XGBoost components, confirming its stability and generalizability.

Keywords: *Fraud Detection, Convolutional Neural Network, XGBoost Algorithm, Apriori Algorithm.*

1. Introduction

The escalating prevalence of fraud poses a significant and costly challenge across various spheres of human life. Individuals and organizations may resort to illicit activities such as deceit or breach of trust for specific gains like financial advantages or other unlawful motives. Traditional fraud detection methods often revolve around the concept of the fraud triangle, comprising motivation, rationalization, and opportunity [26].

* Corresponding Author
[1] Department of Computer Engineering, FSh.C., Islamic Azad University, Fouman, Iran, Email: me.farrokhbakht@iau.ac.ir.
[2] Department of Computer Engineering, FSh.C., Islamic Azad University, Fouman, Iran, Email: ali_akbar_akhavan@yahoo.com.

In recent years, fraud has garnered heightened attention from investors and accountants due to its implications. It not only jeopardizes investors' risks stemming from fraudulent financial practices but also casts doubt on the credibility of accountants. Consequently, accountants strive to fortify financial statement audits by adhering to a set of standards and principles to minimize discrepancies in financial reports. Given the absence of a universal model tailored to individual countries' requirements, this study endeavors to formulate a model based on existing auditing standards as a benchmark for auditors to enhance the detection and mitigation of financial frauds [1, 2].

Fraud manifests in diverse forms, each employing distinct methodologies, with this study focusing specifically on fraud within financial statements. Financial statements encapsulate an organization's financial activities, offering a comprehensive overview of its financial performance from various angles [16]. Key components of these reports encompass expenses, income, loans received or extended, profit, and loss [20]. Among these elements, financial expenses and resultant losses hold paramount importance [20]. Timely identification of such malfeasance can avert substantial financial losses. Historically, conventional fraud detection approaches have centered around the "fraud triangle" model, comprising motivation, rationalization, and opportunity [9, 15].

Individuals and entities may engage in fraudulent practices, including embezzlement, driven by motives like financial gain [23-22]. The American Institute of Certified Public Accountants (AICPA) broadly defines fraud, encompassing a spectrum from minor employee pilferage to elaborate asset misappropriation and falsification of financial statements [17].

The vast array of financial figures presents a ripe opportunity for fraudsters to engage in illicit activities. Common forms of financial statement fraud encompass premature revenue recognition, unrealized gains or losses, asset overstatement, expense understatement, and the concealment or misrepresentation of expenses [6, 7]. According to the Association of Certified Fraud Examiners (ACFE), financial statement fraud ranks as the third most prevalent type of corporate fraud, following corruption and embezzlement. Auditors typically strive to uncover fraudulent activities within financial institutions by scrutinizing motivations behind fraudulent behavior and devising new fraud detection models [25]. However, a 2022 report by the ACFE reveals that internal and external auditors of institutions only identified 16% and 4% of suspicious fraud cases, respectively [10]. This ineffectiveness can be attributed to the rapid evolution of technologies, the dynamic nature of fraud tactics, limited fraud detection patterns, and a lack of expertise in data mining, rendering traditional auditing methods ineffective and obsolete against modern fraud schemes [11, 12].

Recent years have witnessed a surge in the research and implementation of intelligent systems tailored to detect financial statement fraud effectively. These systems offer auditors early warning mechanisms, streamlining decision-making processes [13]. A survey of existing literature underscores the prevalent use of diverse data mining methodologies in fraud detection due to their efficacy in identifying financial statement fraud and other financial irregularities like check fraud, loan fraud, and credit card fraud [27, 5].

Machine learning algorithms have emerged as pivotal tools in data mining, enabling the discovery and exploitation of latent relationships and events within vast datasets. Nevertheless, these approaches encounter challenges such as high dimensionality of data, absence of fraud-ready models, and data imbalance, necessitating meticulous consideration [14-17].

The benefits of employing artificial intelligence in fraud detection are manifold. Noteworthy advantages include heightened accuracy, speed, efficiency, seamless integration with diverse datasets, and cost reduction, stemming from decreased reliance on additional auditors and enhanced fraud detection efficiency.Utilizing artificial intelligence as a sophisticated and potent tool for fraud detection in financial statements not only enhances accuracy and efficiency but also delivers substantial economic and operational advantages to organizations.

In general, data mining techniques for addressing the challenge of fraud detection in financial statements can be categorized into two main types based on the data utilized: binary classification and anomaly detection. In binary classification, statistical models or machine learning methods are employed to classify samples into fraud-prone and fraud-free categories when a sufficient number of fraud-prone samples are available in the dataset. On the other hand, anomaly detection not only identifies fraud but also pinpoints the type of anomaly present [13].

While traditional machine learning models have laid the groundwork, recent years have witnessed a significant shift towards deep learning architectures for detecting complex, non-linear patterns in financial statement fraud. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTMs) and Bidirectional LSTMs (Bi-LSTMs), have been employed to model the sequential nature of financial data over time, showing improved performance over static models [21]. More recently, Transformer-based models, with their self-attention mechanisms, have demonstrated superior capability in capturing long-range dependencies and contextual relationships within financial text and numerical sequences, outperforming RNN-based approaches in several benchmarking studies [28]. However, a key limitation of these advanced deep learning models is their inherent "black-box" nature and their heavy reliance on vast amounts of labeled data, which is precisely the scarce resource in fraud detection scenarios. This creates a critical gap for methods that can leverage the representational power of deep learning while operating effectively under the severe class imbalance typical of financial fraud datasets.

Detecting fraud within financial statements holds immense importance as financial fraud can trigger direct negative repercussions for individuals, companies, organizations, and even the macroeconomy. Key applications of financial statement fraud detection encompass:

1. Detecting financial document forgery: It is critical to identify fraud in financial documents like pay slips, bank statements, and invoices. This entails detecting forged or manipulated documents used illicitly to secure loans or fabricate misleading financial reports.
2. Analyzing unusual patterns: Leveraging analytical software and AI algorithms to spot irregular patterns in financial data can be instrumental. Unusual transactional or payment patterns may serve as indicators of potential fraud.
3. Verifying document consistency: Scrutinizing financial documents for alignment with other information such as financial reports, tax filings, and bank records is pivotal in fraud detection. This verification process plays a key role in ensuring the accuracy and legitimacy of financial data.
4. Identifying financial report fraud: Fraud within financial reports like income statements, balance sheets, and cash flow statements can have severe repercussions on companies and investors. Employing analytical techniques and meticulous examination of these reports aids in identifying fraudulent activities.

Detecting fraud in financial statements serves as a crucial tool for financial risk management, enabling companies and organizations to mitigate potential financial losses that could negatively impact them. By effectively identifying and preventing financial fraud, organizations can bolster their risk management strategies and safeguard their financial health.

In general, the detection of fraud in financial statements is vital for fostering transparency, building trust, and fortifying economic resilience within any organization or society. Leveraging technology, analytical software, meticulous scrutiny of financial documents and data, as well as field research, can enhance performance in this domain and proactively prevent instances of financial fraud.

Despite the benefits of fraud detection through data mining, several shortcomings persist:

a) **Insufficient accuracy:** Some fraud detection methods in data mining lack the required accuracy, leading to potential misidentifications of fraud and the risk of unfairly targeting innocent individuals.

b) **Insufficient efficiency:** Certain fraud detection methods in data mining are not sufficiently efficient, demanding significant time and financial resources for implementation, hindering their widespread adoption.

c) **Circumvention:** Fraudsters can devise new methods to circumvent existing fraud detection techniques in data mining, undermining the effectiveness of these measures [11].

Addressing these shortcomings is paramount to fortifying fraud detection practices in fields reliant on data mining. To overcome these challenges, further research is essential in the following areas:

- Designing fraud detection methods in data mining with enhanced accuracy and efficiency.
- Identifying and mitigating new methods of circumventing fraud detection techniques in data mining.

In this paper, an intelligent fraud detection method utilizing XGBoost will be introduced. The structure of the paper is organized as follows: an introduction, presentation of the proposed method in the second section, analysis of results in the third section, and a conclusion in the fourth section.

1.1. Research Gap

Despite significant progress in applying artificial intelligence to financial fraud detection, critical research gaps remain that hinder optimal performance in real-world scenarios. First, severe class imbalance in fraud datasets, where fraudulent cases are rare, often degrades model performance, yet few studies effectively integrate advanced data augmentation techniques like Generative Adversarial Networks (GANs) specifically adapted for structured financial data. Second, many existing approaches rely on isolated model architectures, leaving a gap in developing optimized, hybrid pipelines that systematically combine feature selection, intelligent data augmentation, and robust classification tailored for financial statements. Third, while high-dimensional financial data can introduce noise, the use of association rule mining for feature selection within deep learning frameworks remains underexplored. To address these gaps, this paper introduces a novel three-stage pipeline that sequentially integrates the Apriori algorithm for feature selection, a custom GAN with a CNN-based generator for imbalance-aware data augmentation, and XGBoost for final classification. The key contributions and novelty of this work are threefold: (1) the proposed hybrid architecture is specifically designed to handle the challenges of financial data by combining feature selection, augmentation, and classification into a cohesive framework; (2) the use of a CNN-based generator within a GAN to synthesize realistic tabular financial data represents a significant advancement over conventional oversampling methods or image-oriented GANs; and (3) comprehensive empirical validation demonstrates superior performance compared to benchmark models (1D-CNN, LSTM, Bi-LSTM), with additional sensitivity analyses confirming robustness across varying data conditions. Thus, the primary novelty lies in the design and validation of this synergistic pipeline, which

effectively addresses data imbalance, dimensionality, and accuracy requirements in financial statement fraud detection.

2. Materials and Methods

To detect fraud, data is initially sourced from the Kaggle database. Subsequently, feature selection is conducted utilizing the Apriori algorithm. To expand the dataset, data is processed through the GAN-CNN network. Finally, fraud detection is executed using XGBoost. The methodology flow is depicted in Figure (1), illustrating the proposed approach. Detailed elaboration on this method will follow below.

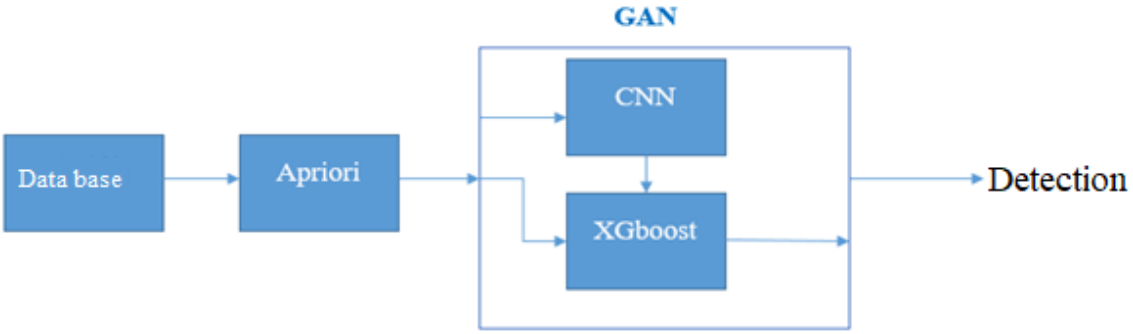

Figure 1. Flowchart of the proposed method

Figure 1, shows the three-stage pipeline of the proposed fraud detection method. The process begins with raw financial data from the Kaggle database. Stage 1 applies the Apriori algorithm for feature selection to reduce dimensionality. Stage 2 uses a GAN with a CNN generator (GAN-CNN) to augment the dataset by generating synthetic samples to address class imbalance. Stage 3 employs the XGBoost classifier to perform the final fraud detection. Arrows indicate the sequential flow of data through the pipeline.

2.1. Apriori Algorithm

Feature selection is a critical pre-processing step [4]. The Apriori algorithm is an algorithm used to find a set of repeated items. Apriori is a surface search algorithm that moves to the next step, k+1, after completing the search in the kth step. This process is repeated until the final condition or conditions are met. In the kth step, a set of k items will be generated. After calculating the support value for each and comparing it with the minsup value (meaning the number of repetitions among several transactions, which is considered 60% here), k repeated patterns are identified. In the next step, the algorithm uses the k repeated patterns to generate a set of (k+1) candidate items that can potentially be repeated. Similarly, according to the minsup value, some are eliminated and a set of (k+1) repeated items will be formed. This process continues until the last repeated item set is found. With the help of this algorithm, the search space is reduced.

2.2. GAN-CNN Network

In the realm of intelligent algorithms and neural networks, sufficient data quantity is crucial for adequately training the model to achieve desired outcomes. To amplify the dataset, a Generative Adversarial Network (GAN) is employed. The GAN comprises two key components: the Generator and the Discriminator, which engage in a competitive process to scrutinize, record, and replicate changes within the dataset.

- **Generator:** The Generator employs a convolutional neural network (CNN) due to its adeptness in deep learning techniques and management of vast data quantities, making it a formidable tool for data generation.
- **Discriminator:** Utilizing the XGBoost algorithm, the Discriminator distinguishes between authentic and generated data samples.

The GAN-CNN network, illustrated in Figure 2, showcases the interplay between the Generator utilizing CNN for data generation and the Discriminator employing XGBoost for fraud detection.

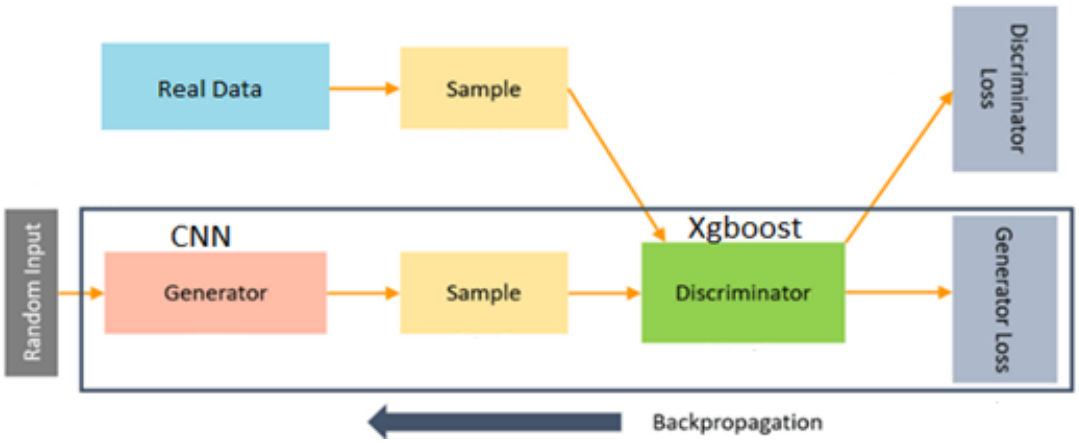

Figure 2. GAN-CNN network

Figure 2, shows the architecture of the GAN-CNN network used for data augmentation. The Generator is a 3D Convolutional Neural Network that creates synthetic financial data samples. The Discriminator is an XGBoost model that evaluates whether an input sample is real (from the original dataset) or fake (generated by the CNN). The two components are trained adversarially: the Generator aims to produce data that fools the Discriminator, while the Discriminator aims to correctly classify real vs. fake data. This competition leads to the generation of high-quality, realistic synthetic fraud samples.

The CNN network's primary advantage lies in its ability to extract essential features from input data without necessitating extensive preprocessing. By adjusting filters concurrently during training, CNN networks yield robust results even with substantial datasets. The convolutional neural network comprises three integral components: the convolutional layer, pooling layer, and fully connected layer, which executes the classification process. After feature extraction and computation, the classification layer assigns random weights to inputs and predicts the appropriate label. Ultimately, the final prediction is derived from the last layer. In the proposed model, input from the classification

layer is channeled to the adversarial module, XGBoost, to execute the classification operation effectively.

The convolutional neural network used in the proposed model is 3D and has 6 hidden layers. The dimensions of each input data are 640×14×1. The convolutional layer operates on the input with a 4×2×20 filter bank and produces an output value with dimensions of 637×13×20C1=. In the pooling layer S2, the filter bank size is 2×2×20 and its output is 319×7×20. In the next convolutional layer, C3, the filter bank kernel is 4×2×20 and its output dimensions are 316×6×20.

The pooling layer in stage S4 has a 2×2×20 kernel and an output with dimensions of 158×3×20. In the last convolution layer C5, a 4×2×1000 kernel will be used, the output of which has dimensions 155×2×1000, and in the last pooling layer S6, a 2×2×1000 kernel will be used, and the output will be 78×1×1000. This value is fed into the Xgboost module. In fact, more data is generated by the generator module using a convolutional neural network, then fraud is detected based on fake and real labeling data using XGboost.

2.3. XGBoost Algorithm

The XGBoost algorithm, a robust tool rooted in decision trees and harnessing gradient boosting techniques, stands out for its exceptional accuracy and efficiency. As a supervised learning algorithm, XGBoost is renowned as one of the most potent and efficient tools in the realm of machine learning. Its prowess lies in self-tuning capabilities, accelerated training speeds, and predictive strength. Widely utilized across diverse domains, from financial data analysis to medical prognostication, XGBoost represents a pinnacle in predictive modeling [4].

The amalgamation of decision trees and gradient boosting has elevated XGBoost to unmatched performance levels. Leveraging sophisticated techniques, this algorithm has redefined the efficacy of decision trees. By employing objective functions and intricate optimization methodologies, XGBoost delivers precise and efficient predictions surpassing the capabilities of conventional decision trees.

In practice, XGBoost has demonstrated its prowess in identifying fraudulent financial transactions. By discerning anomalous patterns in customer behavior, this algorithm plays a pivotal role in thwarting fraudulent activities. Figure (3) provides a detailed insight into the fraud detection process facilitated by XGBoost.

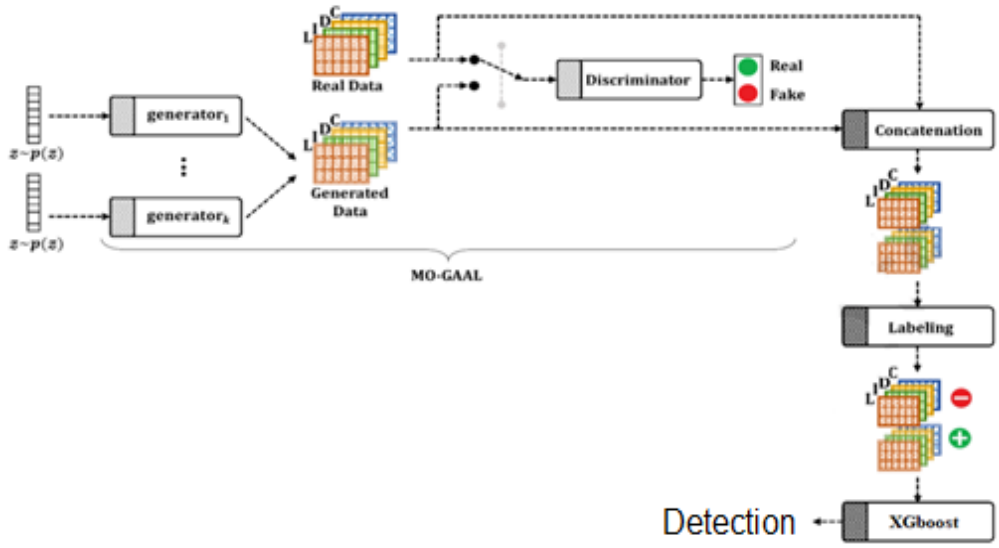

Figure 3. Fraud detection process using the proposed method

Figure 3, detailed workflow of the fraud classification process using XGBoost. After data augmentation via GAN-CNN, the processed dataset (containing both original and synthetic samples) is fed into the XGBoost classifier. The diagram illustrates the internal mechanism: (1) Input features are passed through an ensemble of sequentially built decision trees, (2) Each tree contributes a prediction, and (3) The final fraud/non-fraud classification is obtained by aggregating (boosting) the outputs of all trees. The result is a highly accurate and interpretable fraud risk score for each financial statement.

3. Experimental Results

The evaluation of AI models involves scrutinizing and assessing their performance and efficiency. This critical process aims to gauge the capabilities, effectiveness, and reliability of AI models across diverse applications. In this study, metrics such as precision, accuracy, recall, and F-score have been employed to evaluate the simulated outcomes [3].

Where:

• TP: The algorithm classified the sample in the positive category and the sample is positive.

• FP: The algorithm classified the sample in the positive category but the sample is negative.

• TN: The algorithm classified the sample in the negative category and the sample is negative.

• FN: The algorithm classified the sample in the negative category but the sample is positive.

The Python programming language has been used to implement the proposed method. The k-fold method, k=10, has been used to train and validate the proposed method.

The results of evaluating the proposed method with different criteria are shown in Table 1.

Table 1. Evaluation of the proposed method with different criteria

	Accuracy	Precision	Recall	F-measure
Proposed Method	96.78	95.33	94.13	93.99

In this section, different methods for detecting fraud in financial statements are compared on the used dataset. To examine the generalizability, the proposed method is first compared with different methods. Table 2 shows the detection accuracy.

Table 2. Accuracy of different methods

Method	Accuracy	Recall	Precision
1D-CNN	90.31	91.8	89.4
LSTM	87.45	85.9	86.35
Bi-LSTM	93.31	92.98	93.02
Proposed Model	96.21	95.41	96.02

Table 2 highlights the superiority of the proposed method over other approaches, showcasing a higher detection accuracy. Subsequently, the accuracy, precision and recall of the proposed method has been compared with alternative methods using varying amounts of data. Figure 4, 5 and 6 illustrate the comparative evaluations.

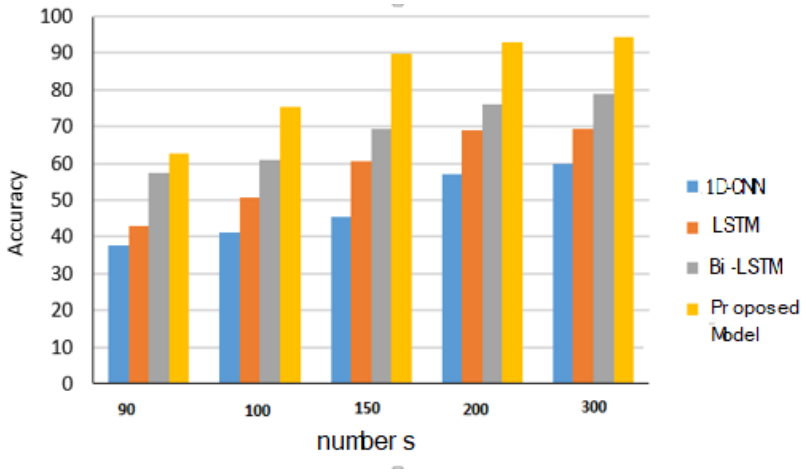

Figure 4: Evaluation of the accuracy of the some methods with different data numbers

In Figure 4, the plot compares the proposed method against three benchmarks (1D-CNN, LSTM, Bi-LSTM). The proposed method (solid blue line) consistently achieves the highest accuracy across all data sizes and demonstrates the least performance degradation when less data is available, highlighting its robustness and data efficiency.

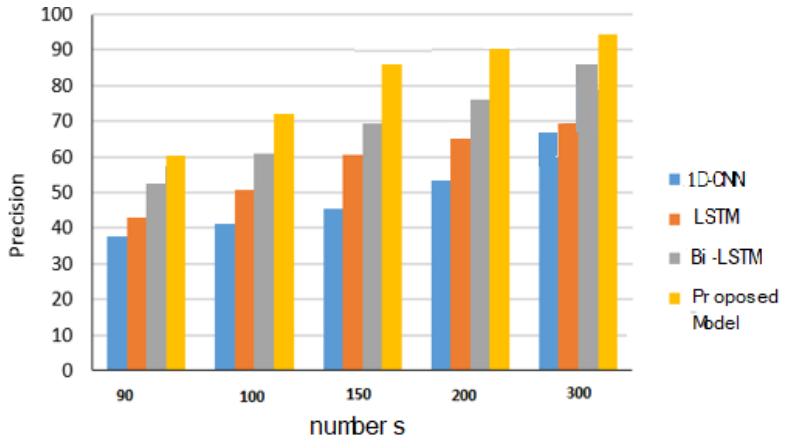

Figure 5: Evaluation of the precision of the some methods with different data numbers

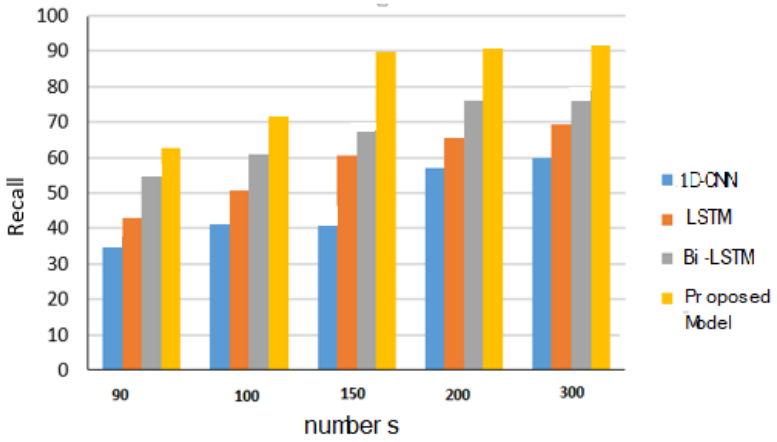

Figure 6: Evaluation of the recall of the some methods with different data numbers

As can be seen in the figures above, the proposed method performs detection with higher accuracy, precision and recall in different numbers of data. In second place, Bi-LSTM has higher performance. Also, different amounts of data affect efficiency.

4. Conclusion

This paper introduced an intelligent hybrid approach for detecting fraud in financial statements. The proposed method integrates feature selection using the Apriori algorithm, data augmentation via a GAN-CNN network, and final classification with XGBoost. The experimental results unequivocally demonstrate the superior performance of the proposed method. It achieved an accuracy of 96.21%, precision of 96.02%, recall of 95.41%, and an F1-score of 93.99% on the benchmark dataset. This represents a significant improvement over the compared baseline models (1D-CNN, LSTM, and Bi-LSTM). The synergy between the components is critical: the Apriori algorithm effectively reduced dimensionality and noise, the GAN-CNN successfully mitigated the class imbalance problem by generating realistic synthetic samples, and the XGBoost classifier provided robust and interpretable final detection. The sensitivity analysis further confirmed the model's stability across different data sizes.

Despite its promising results, this study has certain limitations. First, the model was trained and validated on a specific, publicly available dataset from Kaggle. Its performance on proprietary, real-time, or industry-specific financial data streams requires further validation. Second, while the GAN-CNN augmentation improves performance on imbalanced data, the quality of generated samples depends heavily on the initial data distribution and GAN training stability, which can be challenging. Third, the "black-box" nature of deep learning components (GAN, CNN) somewhat limits the interpretability of the feature generation process, even though XGBoost offers some feature importance insights.

Future research can build upon this work in several directions such as Domain Adaptation for testing and adapting the model to diverse financial domains and Real-time Detection for developing a framework for near real-time fraud detection by optimizing the pipeline for streaming financial data. The findings of this study offer actionable insights for practitioners in audit, compliance, and financial governance. The high accuracy and robustness demonstrated by the proposed model suggest it can be effectively deployed as a pre-screening tool to prioritize high-risk financial statements, thereby

optimizing audit resource allocation. Furthermore, the model's stability across varying data volumes, as evidenced by the sensitivity analysis, indicates that organizations can implement a functional version even with limited historical fraud data, allowing for incremental improvement as more data becomes available. Additionally, the integration of interpretable machine learning (XGBoost) enables auditors to identify key fraud indicators, supporting not only detection but also the design of targeted internal controls. These insights collectively underscore the practical viability of adopting hybrid AI systems to strengthen fraud detection frameworks, reduce operational risks, and enhance trust in financial reporting.

References

[1]	Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*, 100402.
[2]	Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. *Ieee Access*, *10*, 72504-72525.
[3]	Askari, E., Setarehdan, S. K., Sheikhani, A., Mohammadi, M. R., & Teshnehlab, M. (2018). Modeling the connections of brain regions in children with autism using cellular neural networks and electroencephalography analysis. *Artificial intelligence in medicine*, *89*, 40-50.
[4]	Bagheri, S., Askari, E., Motamed, S. (2025). Classification of Leukemia Using a Hybrid Approach Based on Temporal Fusion Transformer and XG-Boost. *Iranian Journal of Operations Research*, 16(2), 45-62.
[5]	Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
[6]	Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, *557*, 317-331.
[7]	Chen, J. I. Z., & Lai, K. L. (2021). Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence*, *3*(02), 101-112.
[8]	Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.
[9]	Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*, 113421.
[10]	De Rossi, G., Kolodziej, J., & Brar, G. (2020). A recommender system for active stock selection. *Computational Management Science*, *17*(4), 517-547.
[11]	El Kafhali, S., & Tayebi, M. (2022, December). Generative adversarial neural networks based oversampling technique for imbalanced credit card dataset. In *2022 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)* (pp. 1-5). IEEE.
[12]	Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448-455.
[13]	Gangwar, A. K., & Ravi, V. (2019, December). Wip: Generative adversarial network for oversampling data in credit card fraud detection. In *International Conference on Information Systems Security* (pp. 123-134). Cham: Springer International Publishing.
[14]	Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144.
[15]	Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, *15*(4), 357-380.
[16]	Gupta, A., Vatsa, M., Kumar, V. (2015). A Survey of Data Mining Techniques for Fraud Detection, *ACM Computing Surveys*, 47(4).

[17]	Gupta, S., & Mehta, S. K. (2024). Data mining-based financial statement fraud detection: Systematic literature review and meta-analysis to estimate data sample mapping of fraudulent companies against non-fraudulent companies. *Global Business Review*, *25*(5), 1290-1313.
[18]	Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods. *Knowledge-Based Systems*, *128*, 139-152.
[19]	Hajek, P. (2019, May). Interpretable fuzzy rule-based systems for detecting financial statement fraud. In *IFIP international conference on artificial intelligence applications and innovations* (pp. 425-436). Cham: Springer International Publishing.
[20]	Hashim, H. A., Salleh, Z., Shuhaimi, I., & Ismail, N. A. N. (2020). The risk of financial fraud: a management perspective. *Journal of Financial Crime*, *27*(4), 1143-1159.
[21]	Huang, Y., Li, S., & Wang, F. (2023) A sequential deep learning model with attention for financial statement fraud detection. *Expert Systems with Applications*, 214, 119123.
[22]	Omidi, M., Min, Q., Moradinaftchali, V., & Piri, M. (2019). The efficacy of predictive methods in financial statement fraud. *Discrete Dynamics in Nature and Society*, *2019*(1), 4989140.
[23]	Petković, Z., Milojević, S., Novaković, S., & Trivunović Sajić, Đ. (2021). Fraudulent financial reporting from the managers' perspective. *International Academic Journal*, *2*(2), 35-39.
[24]	Saia, R., & Carta, S. (2019). Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks. *Future Generation Computer Systems*, *93*, 18-32.
[25]	S., Widup, A., Pinto, C.D., Hylender, G., Bassett. (2021) Verizon Data Breach Investigations Report, Report published in resaechgate.
[26]	Shahriari, M., Eshaghinia, M., & Fathihafashjani, K. (2024). Impact of Foreign Investment Risk Factors on Attracting Foreign Investment in Upstream Industries. *Iranian Journal of Operations Research*, *15*(1), 91-77.
[27]	Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision support systems*, *50*(2), 491-500.
[28]	Wang, J., & Liu, H. (2024). FinBERT: A pre-trained transformer model for financial text mining and its application to fraud detection. *IEEE Access*, 12, 24567-24578.