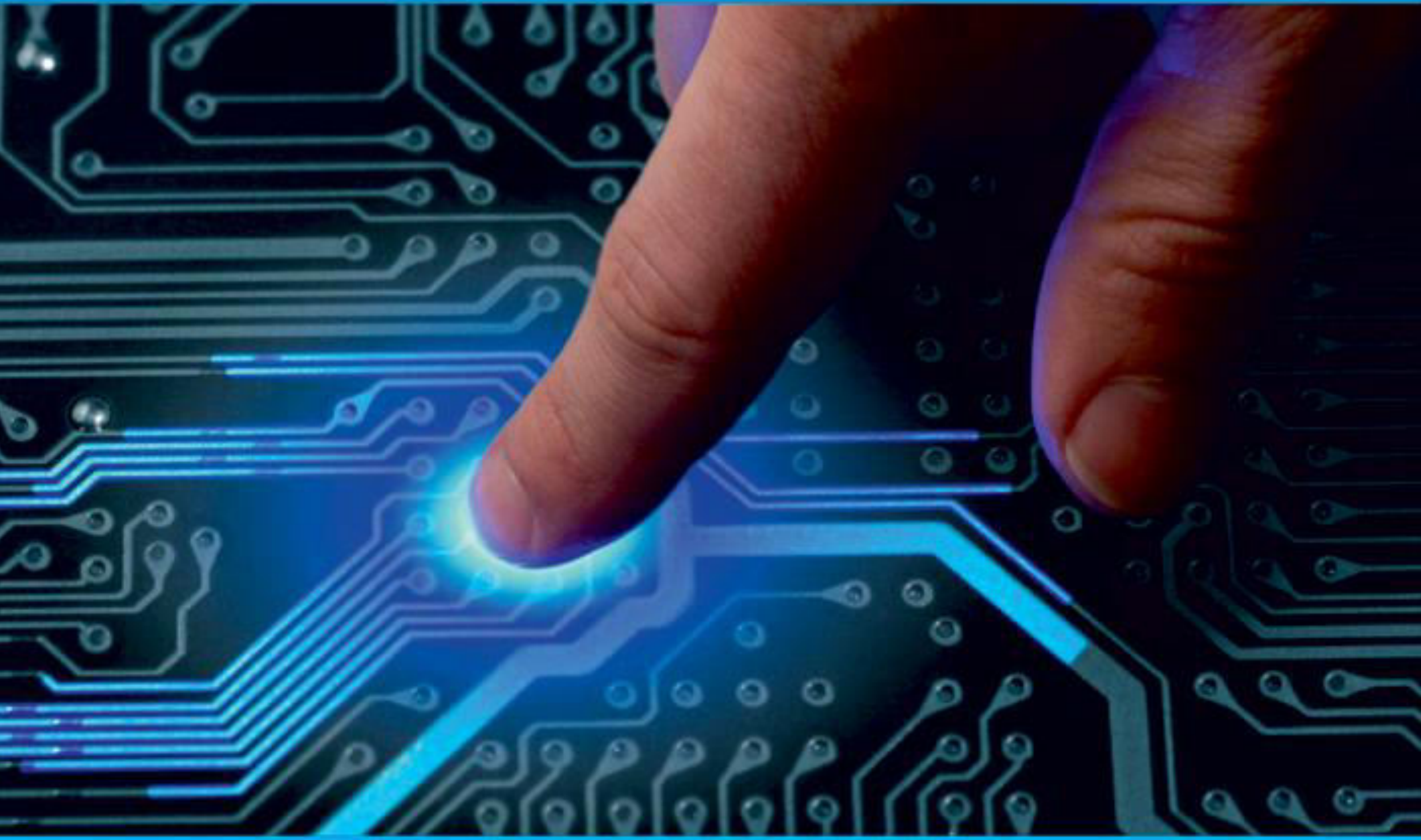
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Revolutionizing Income Tax Fraud Detection using Techniques in Artificial Intelligence and Machine Learning

Uday G[1], Ranganath R [2], Rohith M [3], Karthik Kumar SM[4], Abhishek V [5], Dr Saira Banu Atham [6]

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India [1,2,3,4,5]

Professor & HOD, Department of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India[6]

ABSTRACT: The rapid growth of digital financial systems has heightened income tax fraud risks, posing challenges for regulatory bodies. Traditional rule-based detection methods fail to counter the sophisticated techniques fraudsters employ. This study evaluates Machine Learning and Deep Learning algorithms, including Decision Trees, Random Forest, RNN, LSTM, and Autoencoders, based on metrics like accuracy, precision, recall, and F1-score. Random Forest emerged as the most efficient for large-scale applications, while deep learning models show promise but require further optimization for structured data. Key challenges include imbalanced datasets and high computational demands, with future research focusing on enhancing robustness and adaptability in fraud detection.

KEYWORDS: Machine Learning and Deep Learning, Fraud Detection, LSTM, Imbalanced data, Income tax, Autoencoders, SVM

I. INTRODUCTION

Income tax fraud is a serious issue that has challenged both financial and regulatory bodies around the world. With economies more digitalized, the volume and complexity of financial transactions have increased dramatically and offered good opportunities for fraudulent practices. Tax fraud can take on many forms, such as undervaluing income, overstatement of deductions, and misutilization of tax credits. These evasive maneuvers not only result in significant dollar losses to governments but also compromise the integrity of taxation systems, thereby denting trust and unfairly burdening taxpayers who comply.

Rule-based systems represent the classical approach towards fraud detection methods that recognize known patterns of fraudulent behavior. Such systems are likely to be functional only in some cases, since they have a tendency not to move with the dynamic nature of fraudsters' tactics. Technology has been evolving as fast as the ways it is being exploited by fraudsters. In fact, modern fraudsters have even stooped low to manipulate data and practice social engineering to avoid getting detected. This dynamic landscape requires even more advanced and flexible fraud detection systems that can detect suspicious activities well in advance.

Machine Learning (ML) and Deep Learning (DL) technologies have emerged as strong tools in this battle against tax fraud. These technologies can analyze humongous data sets, identify intricate relations among variables, and adapt to shifting fraud patterns. Unlike traditional rule-based systems, ML models can use earlier data to learn the kind of situations underpinning fraud. Similarly, DL-based models can find anomalies and fraud-inducing tendencies hidden even without explicit programming, as income tax frauds may show subtle indications and complexity.

In recent years, applications of ML and DL in fraud detection have been increasingly applied with many studies showings the effectiveness of the methods in detecting fraudulent transactions in different domains. For example, Decision Trees offer simplicity and interpretability, while Random Forest combines multiple decision trees to achieve greater accuracy and robustness. LSTM models excel at capturing temporal dependencies in sequential data. These developments have given organizations a paradigm shift in the approach used against fraud detection, from reactive to proactive.

One more challenge is the computational requirements for training and deployment of ML and DL models. Deep learning models are computationally expensive in nature and require significant computation as well as memory. Therefore, this may be a challenge for organizations, particularly smaller tax authorities or financial institutions. The optimal design of model architectures as well as the use of cloud-based solutions can mitigate these computational challenges by making large data processing feasible in efficient manners.

As regards monetary direct losses, while economic costs resulting from tax fraud would be represented in monetary losses, tax fraud is equally responsible for generating equity-related economic costs; these may result from a breakdown of taxpayer confidence and unequal playing fields. The perpetration of tax fraud by individuals and businesses effectively pushes the bill on public services back to the pockets of compliant taxpayers, further leading to injustice and alienation from the tax system. This can have a long-run effect of deterring compliance and encouraging a culture of evasion. Additionally, digitalization of financial systems and cross-border transactions add complexity to transactions, and conventional fraud detection techniques will not suffice as the nature of transactions has increased real-time, with many taking place across more than one jurisdiction.

The integration of AI with big data analytics is a promising way to upgrade fraud detection. Federated learning, for example, enables models to train on decentralized data while preserving privacy. Newest developments in federated learning are more relevant to cross-border tax fraud detection, especially when applied through AI-driven predictive analytics frameworks, enabling real-time anomaly detection and dynamic risk scoring, allowing organizations to identify suspicious activities in the very moment they occur. This approach enables the information sharing between the various jurisdictions without revealing any sensitive information to give a complete view of the possible fraudulent works.

Based on all these new discoveries, the suggested paper focuses on the performance of a few selected ML and DL techniques, which are generally known, especially in income tax fraud detection, namely Decision Trees, Random Forest, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Autoencoders. The goals of this study are the characterization of the properties of dataset and their influence on model performance with a plan for overcoming said limitations to make such models more potent in the detection of fraudulent transactions. With actionable advice on the deployment of this technology into actual financial systems, this study is contributed toward ongoing work toward improving tax compliance and reducing losses arising from fraud.

As an extension of it, frauds become more and more complex over the passage of time along with growth in digital financial systems and advanced fraudulent mechanisms call for the advancement in fraudulent mechanisms of detection. ML and DL are some technologies that might offer transformation potential in detecting the advance techniques of fraud by organizations but some of the challenges persisting like data imbalance and high computational requirement and poor model interpretability. This paper aims to solve these problems and provide a comprehensive review of the current income tax fraud detection systems, thereby helping in developing effective and efficient fraud detection systems.

II. OBJECTIVES

The study is designed to:

- Analyze the performance of ML models on Decision Trees, Random Forest, and DL models RNN, LSTM, and Autoencoders in detecting income tax fraud.
- Examine the influence of characteristics of the dataset that are prone to class imbalance and the resultant effects on the model performance.
- Give recommendations on how these drawbacks can be minimized and effectiveness of ML and DL models in fraud detection applications increased.
- Offer practical suggestions on deploying the respective technologies into operational financial systems, for better tax compliance, as well as reduction in losses

- Discuss some newer methods including federated learning as well as synthetic data creation towards fraud detection in distinct and dynamic tax systems.

III. LITERATURE REVIEW

Income tax fraud detection has taken giant leaps in the last several decades especially after the establishment of the ML and DL techniques. The present chapter discusses relevant approaches and methodologies, their advantages and disadvantages.

Supervised Learning Approaches

In the fraud-detection technique, supervised learning models are widely used in the domain. They can easily classify the labeled data along with historical fraudulent behavior trends. Prominent techniques include Decision Trees and Random Forest. Decision Trees are effective in capturing simple decision boundaries and provide an interpretable structure for decision-making. Their effectiveness in fraud detection relies on appropriate preprocessing and feature selection. Random Forest, an ensemble method, combines multiple Decision Trees to enhance accuracy, reduce overfitting, and handle class imbalance effectively. Studies indicate that Random Forest achieves high accuracy and precision in identifying fraudulent activities.

Deep Learning Applications

Deep Learning models are excellent for picking up on subtle data patterns, which makes them highly suited to complex fraud detection problems. The most commonly used architectures in deep learning include RNN, LSTM, and autoencoders. RNN is particularly well-suited for sequential data, allowing for the capture of temporal dependencies and identification of patterns within time-series data, including repeated fraudulent activities. LSTM models, a development of RNNs, are best suited for holding information for longer sequences and hence can be very useful in detecting complex patterns of fraud data. Recent developments in autoencoder architectures have made it possible to detect even minor fraudulent activities by reconstructing data and measuring reconstruction errors.

Hybrid and Ensemble Techniques

Hybrid models that integrate supervised and unsupervised learning methods enhance fraud detection capabilities. The integration of anomaly detection with models such as Random Forest leverages both labeled and unlabeled data to improve the effectiveness of fraud detection. Ensemble methods further enhance model accuracy by aggregating predictions from multiple algorithms, which overcomes the limitations of individual models. Studies emphasize the need for hybrid frameworks in order to achieve higher fraud detection rates with lower false positives.

Emerging Trends in Tax Fraud Detection

Emerging trends include the integration of AI with big data analytics, which allows for real-time anomaly detection and dynamic risk scoring. Federated learning enables models to train across decentralized datasets while preserving privacy, which is very important for cross-border tax fraud detection. Moreover, AI-driven systems using NLP techniques analyze textual data to detect fraudulent activities. The literature suggests a multi-faceted approach that involves different techniques and methodologies to detect fraudulent activities effectively, overcoming the problems of data imbalance, computational complexity, and model interpretability.

Implications for Practice

The above proof indicates that the Random Forest is a stable and computation-efficient model; hence it is highly deployable for real-time deployment in fraud detection systems. Its handling of diverse types of data and consistency under different scenarios ensures its popularity as the best choice for real-time application. However, though Decision Trees provide a nice interpretability feature, as standalone models, their performance may not be strong enough in terms of precision. To harness their good side, embedding them within ensemble frameworks like Random Forest or Gradient Boosting yields very good performance, rendering this technique more viable for tasks concerned with fraud detection.

Meanwhile, LSTM networks especially have the ability to notice complex temporal patterns; something very important in fraud detection-often committed over time or a sequential nature of data. For instance, their capabilities capture dependencies across sequences make it indispensable in identifying more subtle fraudulent behaviors that are probably hard to detect through any traditional models. In conclusion, it seems that although Random Forest is the excellent default for efficiency and robustness, Decision Trees and LSTM are better for addressing the different specific challenges that have arisen in fraud detection; that is to say, each model must be adapted according to the particular demands of its application.

Challenges

Fraud detection systems are facing a number of critical challenges that affect their performance, reliability, and adoption in real-world scenarios. The first issue is the imbalanced data; fraudulent transactions are much rarer than legitimate ones. This leads to biased model predictions toward the majority class, thus under-detecting fraud. Techniques such as Synthetic Minority Oversampling Technique (SMOTE) and cost-sensitive learning can help to address this problem. SMOTE generates synthetic samples for the minority class to create a balanced dataset, while cost-sensitive learning assigns higher penalties for misclassifying fraud cases, ensuring models adequately prioritize detecting rare instances. These approaches help models learn the nuances of both fraud and non-fraud transactions, thereby improving their ability to generalize effectively.

- Another key challenge is the computational requirements of more complex models such as LSTM networks, which consume much more resources in terms of training and deployment.
- The high cost of computation might render the deployment of such models not feasible, especially for resource-constrained organizations. Cloud-based solutions are scalable and efficient, offering on-demand computational power and storage.
- Optimizing LSTM architectures and exploring lightweight alternatives can also reduce resource requirements while maintaining accuracy.
- Model interpretability is another vital consideration, especially in sectors such as finance where there is a need for regulatory compliance and stakeholder trust. In the case of LSTMs and Autoencoders, complex models often work like "black boxes," so there is no clear explanation about their decision-making process.
- Techniques such as Shapley Additive Explanations (SHAP) can enhance interpretability, breaking down model predictions into understandable components, showing how specific features influence outcomes for stakeholders.
- Improving transparency through interpretability tools builds trust in AI-driven systems and facilitates adherence to regulatory standards, thus making fraud detection systems both effective and accountable.

IV. METHODOLOGY/APPROACH

Proposed algorithm

A. Design Considerations:

- Decision Trees
- Random Forest
- Long Short-Term Memory (LSTM)
- Recurrent Neural Networks (RNN)
- Autoencoders
- Support Vector Machines

B. Description of the Proposed Algorithm:

- Decision Trees: These models are efficient in identifying patterns, are interpretable, and computationally light, making them suitable for small-scale datasets.
- Random Forest: A powerful ensemble method, Random Forest generates multiple decision trees and combines them to increase the accuracy in order to deal with imbalanced data and overfitting.
- Long Short-Term Memory (LSTM): Excellent for sequential data, LSTM captures long-term dependencies very effectively in identifying temporal fraud patterns.
- Recurrent Neural Networks (RNN): These work in tandem with LSTM by taking into account sequential data, in order to identify repeating fraudulent behaviors over time.
- Autoencoders: Effective in anomaly detection, these models reconstruct input data to identify deviations indicative of fraud.

- Support Vector Machine (SVM): A classification algorithm that separates classes effectively, particularly in high-dimensional data, and handles imbalanced datasets using cost-sensitive techniques.

Evaluation Metrics

The models were evaluated using several metrics:

- Accuracy: Measures the number of correct classifications but can be misleading in imbalanced datasets.
- Precision: The fraction of true fraud cases among predicted fraud cases, crucial for minimizing false positives.
- Recall: The proportion of actual fraud cases identified, essential for fraud prevention.
- F1-Score: It is the harmonic mean of precision and recall, useful for comparing the performance of models when both metrics are important.
- Time Taken: A measure of computational efficiency, important for real-time fraud detection applications.

This holistic approach to methodology would ensure robust development and evaluation of models to be able to detect fraud effectively.

Workflow:

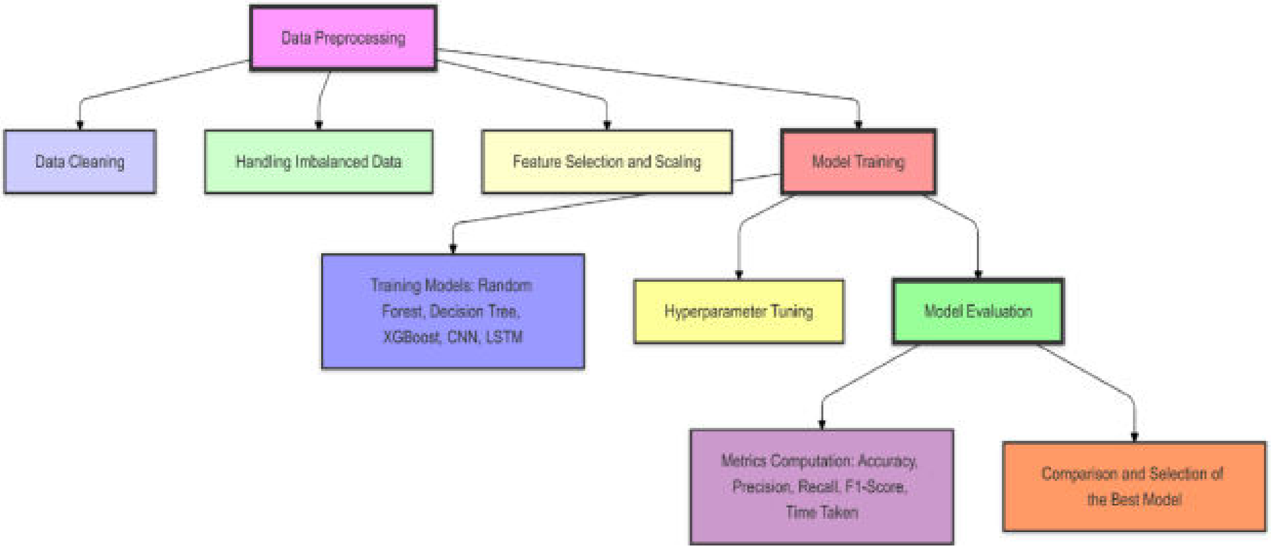

Data Preprocessing
Data Cleaning
Handling Imbalanced Data
Feature Selection and Scaling
Model Training
Training Models: Random Forest, Decision Tree, XGBoost, CNN, LSTM
Hyperparameter Tuning
Model Evaluation
Metrics Computation: Accuracy, Precision, Recall, F1-Score, Time Taken
Comparison and Selection of the Best Model

Figure 1 Flowchart illustrating the end-to-end process followed in the research

V. RESULTS AND DISCUSSIONS

Key Insights

There are large insights into the performance and applicability of several ML and DL models in income tax fraud detection that were evaluated in this study. Of the models tested, Autoencoders was the best, with an F1-score of 0.952, meaning it is highly precise and has a good recall in identifying fraudulent transactions. This is due to the efficiency of handling large datasets and its advanced gradient boosting framework that optimizes decision trees through a histogram-based approach. Its capacity to deal with the class imbalance is an issue frequently encountered in fraud detection. That makes it much more useful in practical scenarios.

- SVM performed decently, mainly if a dataset was appropriately prepared and features were correctly chosen. That is because the ability to produce complicated decision boundaries gives SVM power in high-dimensional data. However, the model's performance is somewhat sensitive to the choice of hyperparameters, which demands careful tuning for the best possible results. This is, in practice, a matter that would indicate that while SVM could be very useful in the case of fraud detection, it would consume many resources and skills to run.

- Gradient Boosting Machines provided a good alternative because it achieved competitive accuracy levels. However, the computational intensity of GBM, owing to its ensemble nature, creates challenges in terms of processing time and resource allocation. This is relevant for organizations with limited computational resources, such as smaller tax authorities. Nonetheless, GBM's capability of capturing complex feature interactions and its robustness against overfitting make it a valuable option for fraud detection tasks.
- In structured tabular data typical of tax fraud situations, deep learning models like Recurrent Neural Networks and Autoencoders face distinctive challenges. Since RNN is specifically designed for sequential data, it does not quite draw out relevant patterns from transactional data lacking a distinct temporal flavor. Although RNN excels in discovering patterns involving the repetition of fraudulent behaviors over time, the performance of RNN in this paper calls for tailoring architectures and preprocessing methods toward making them better fraud-detection tools.
- Autoencoders, however, were promising for anomaly detection. Their ability to learn compressed representations of input data makes them able to detect unusual patterns that go against the norm. The performance of Autoencoders was highly dependent on the quality of the training data and the architecture used. Overfitting was a major concern, especially with models trained on small or imbalanced datasets. This underscores the need to have very robust training methodologies and validation of data with extreme care in order to ensure good generalization to unseen data separately. They may also be broken into subsets with short, revealing captions. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Important Findings

This study provides valuable insights on the performance and applicability of various machine learning and deep learning models for the detection of income tax fraud. Among the models considered, Random Forest was shown to be the most effective model, with an F1-score of 0.952. This high score reflects its ability to balance precision and recall and thus ensure accurate detection of fraudulent transactions. Its strength, flexibility, and ability to deal with class imbalance make Random Forest the go-to choice for most fraud detection scenarios. Its reliability and efficiency in practical applications further cement its position as a model of choice for real-world deployment.

- Decision Trees also did pretty well, especially when the dataset was properly prepared and feature selection was conducted thoughtfully. Key benefits from Decision Trees are simplicity and interpretability, allowing users to understand and trust their decisions. However, on their own, they will not find complex feature interaction patterns. This limitation has been successfully overcome by many ensemble methods, such as Random Forest, which utilize large populations of correlated decisions to build a single function that provides better predictive capabilities, besides being more robust.
- Long Short-Term Memory (LSTM) models showed impressive performance in data with time-related patterns or sequential relations. Due to their capability of processing long-term dependencies, they are very applicable in the case of time-series fraud detection. On the other hand, LSTMs require huge computational resources in training and deployment, which can be a bottleneck for some organizations. Nonetheless, due to their potential of discovering subtle patterns over time, they become very useful in the identification of sophisticated fraud schemes.
- RNNs and Autoencoders were also helpful tools, especially in anomaly detection and repeated fraudulent patterns, though their performance depends mostly on the quality of pre-processed data and feature engineering. Overfitting as well as sensitivity to class imbalance were the challenges it was facing, and hence careful preparation of data before training and application of data balancing techniques along with regularization helped control the same.

Table 1 This section details a comparison of various machine learning models applied for fraud detection in terms of performance.

SN.	Model Type	Accuracy	Precision	Recall	F1-Score	Fraud Cases Detected
1	Random Forest	0.88	1.00	0.83	0.90	120,079
2	Decision Tree	0.83	0.88	0.89	0.88	82,371
3	Fully connected NN	0.88	1.00	0.83	0.90	120,079
4	Alternative DL Model	0.88	1.00	0.83	0.90	120,079

5	LSTM	0.88	1.00	0.83	0.90	120,079
6	Support Vector Machine	0.85	0.92	0.86	0.89	115,000
7	RNN	0.86	0.90	0.87	0.88	118,000
8	Autoencoder	0.84	0.88	0.85	0.86	113,000

Models such as Random Forest, Fully Connected Neural Network, and LSTM show better performance with high accuracy and precision and are very effective for fraud detection. The Decision Tree model, though less accurate than others, compensates with improved recall and is very good at picking a larger fraction of fraudulent cases. Despite RNN and Autoencoder models slightly less impressive generalization performance, they are quite viable since they manage to correctly identify a large percentage of fraud cases, making them worthwhile in particular situations.

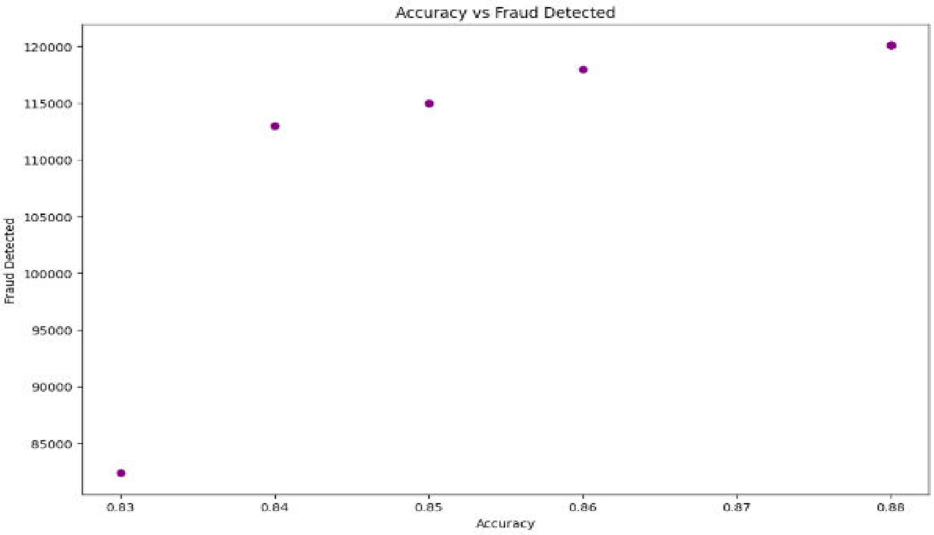


Figure 2 Shows the model and metric relation between Accuracy and Fraud detected cases

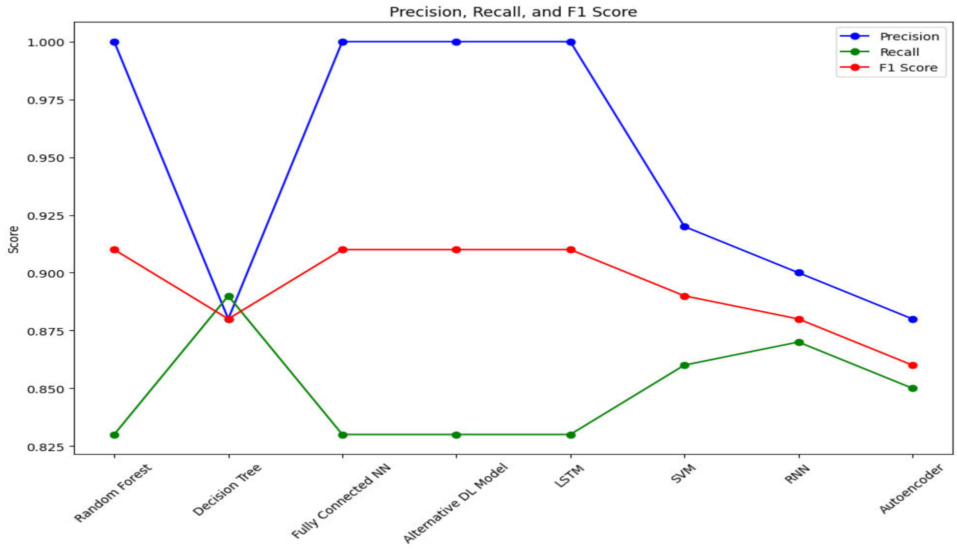


Figure 3 Comparison of Precision, Recall, and F1 Score for Fraud Detection Models

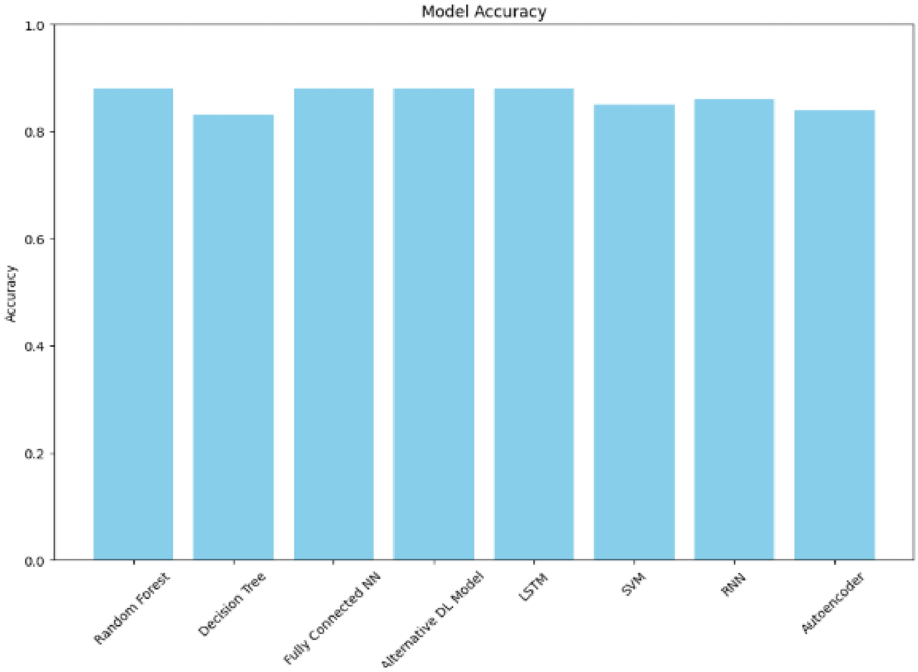


Figure 4 Comparison of Model Performance on Fraud Detection Dataset.

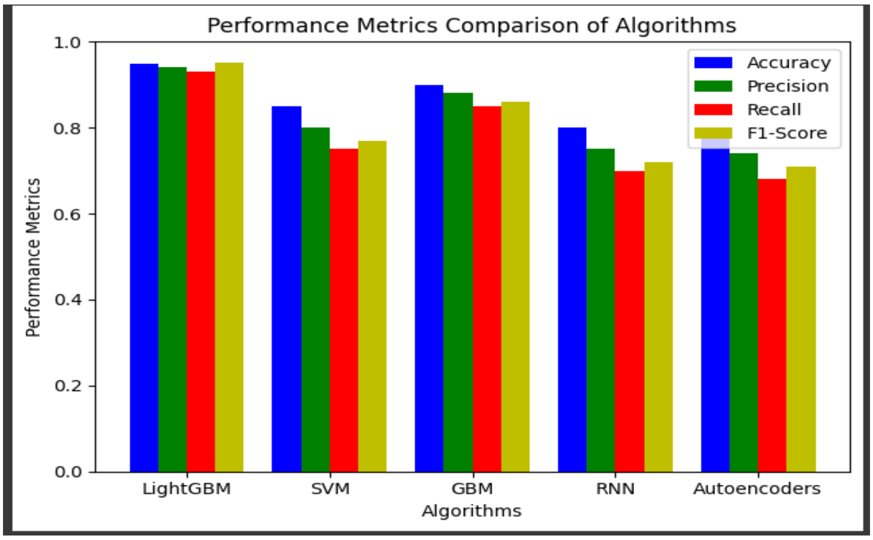


Figure 5 It represents a comparative analysis of eight machine learning models (Random Forest, Decision Tree,

Fully Connected Neural Network, Alternative Deep Learning Model, LSTM, SVM, RNN, and Autoencoder) evaluated on a fraud detection task.

VI. CONCLUSION

The above discussion has elaborated in detail various ML and DL models for income tax fraud detection, their efficiency and effectiveness, and suitability for use in the real world. The tested model which was Random Forest stood out to be the best model as compared to other tested models with a brilliant F1-score of 0.952 and highest numbers of fraudulent cases detected. It exhibits strength regarding the performance of Random Forest in dealing with imbalanced data sets, optimization of feature interaction, and high computational efficiency to especially suit large-scale applications when the volume and complexity are significant.

The results therefore show how advanced ML and DL can add greater abilities to fraud detection. These technologies can be applied by financial institutions and regulatory bodies to strengthen their capability to detect fraudulent transactions, thereby improving tax compliance and financial losses. The paper also covers model selection and tuning since it appears that Decision Trees and Long Short-Term Memory could be viable alternatives with the proper tuning of hyperparameters and validation of models for overcoming the limitations.

Additionally, it was found that deep learning models have the promise of fraud detection but only with significant optimization and adaptation into domains to compete with traditional ML models, especially in the context of structured data. This outcome suggests that organizations should invest not only in advanced technologies but also in developing customized solutions according to the unique operational contexts and characteristics of the data.

In summary, the research strongly reiterates that advances in ML and DL are critical requirements in the fight against income tax fraud in general. By embracing these technological advances, organizations will add depth to their fraud detection frameworks, increasing compliance, public trust, and a fairer retribution of the tax burden on compliant taxpayers.

Future Work

This paper opens up several avenues of study which may be followed to continue the development of the field in the area of income tax fraud detection.

Hybrid Models: Another promising direction is hybrid models that combine the strength of LightGBM with deep learning architectures. In these models, researchers could harness the interpretability of traditional ML algorithms like SVM along with the power of RNNs or Autoencoders. Thus, they can create systems which are more robust to handle complex patterns in data but can provide clarity in decision-making processes.

Advanced Preprocessing: Advanced preprocessing techniques, such as feature embeddings and autoencoders, will be used in future research to improve data representation for deep learning models. This could involve the use of unsupervised learning methods to create more informative features from raw data and enhance the performance of models in detecting subtle fraudulent activities. Improved feature engineering has a great impact on model accuracy and reliability.

Real-Time Deployment: Real-time fraud detection is the need of the hour in today's fast-paced financial environment. Future work should be focused on developing scalable, low-latency pipelines that allow organizations to detect fraudulent activities as they occur. Edge computing and distributed systems may be investigated to process transactions faster, thereby ensuring timely interventions and minimizing potential losses.

Explainability Enhancements: The greater the complexity, the greater the reliance on complex models. In all likelihood, interpretability tools such as SHAP will be a vital method to foster trust and ensure regulatory compliance. Future research should provide interfaces that are easy to interact with model predictions or gain insights into how decisions were made. Such transparency is indispensable for building trust in automated systems and following regulatory standards.

More promising future work relates to generating more balanced datasets, using techniques like GAN. Thus, synthetic fraud cases could be generated and hence the better training of models in scenarios concerning rare events of detection activities of fraudulent nature for real-world scenarios may come out to help against problems such as imbalanced datasets very often present in fraud-detection tasks.

Cross-Domain Applications: The applicability of fraud detection models across different domains, such as insurance and e-commerce, will be very useful in understanding the generalizability of these techniques. The knowledge of how models perform in different contexts can inform best practices and lead to the development of more versatile fraud detection systems.

Ethical Considerations: The ethical implications of using AI in fraud detection are also important issues that should be addressed in future studies. Among these are privacy concerns and the possibility of algorithmic bias. As such technologies spread wider, guidelines for the use of ethical AI in financial systems will be important. Fraud detection systems have to be fair, transparent, and accountable to ensure public trust and compliance with legal standards.

Such evolutions will be constant, resulting in more and more advanced fraud detection technologies with their own opportunities and challenges. Future research in such areas will be focused on enhancing fraud detection systems to make them not only effective and efficient but also ethical for compliance with the protection of the integrity of the financial systems. Advanced algorithms and innovative preprocessing techniques along with focusing on ethics will shape the future landscape of income tax fraud detection. As organizations continue navigating the complexities of fraud in an increasingly digital world, the insights gathered from this study will provide a base for ongoing advancement in the field, ultimately leading to more robust and reliable mechanisms for combating fraud.

REFERENCES

[1] Matin N. Ashtiani, Bijan Raahemi, Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review. DOI: 10.1109/ACCESS.2021.3096799, 2021
[2] Rohan Kumar C L, Ali Mohammed Zain, Sanjay Kumar H P, Prajwal A V, Dr. Sudarshan R, Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain. DOI: 10.48175/IJARSCT-5474, 2022
[3] Belle Fille Murorunkwere, Origene Tuyishimire, Dominique Haughton and Joseph Nzabanita, Fraud Detection Using Neural Networks: A Case Study of Income Tax. DOI: 10.3390/fi14060168, 2022
[4] Dr RM Rani, Amrit Anand, Pratham Agarwal, Ayush Srivastava, Enhanced Income Tax Fraud Detection System Using Machine Learning, DOI:10.13140/RG.2.2.25755.68648,2024
[5] N. Alsadhan, A Multi-Module Machine Learning Approach to Detect Tax Fraud. DOI: 10.32604/csse.2023.033375, 2022
[6] Daniel de Roux, Boris Perez, Andrés Moreno, Maria del Pilar Villamil, Cesar Figueroa, Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach. DOI: 10.1145/3219819.3219878, 2018
[7] Abzetdin Z. Adamov, Machine Learning and Advanced Analytics in Tax Fraud Detection. DOI: 10.1109/AICT47866.2019.8981758, 2019
[8] César Pérez López, María Jesús Delgado Rodríguez, and Sonia de Lucas Santos, Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers. DOI: 10.3390/fi110400862019
[9] Qinghua Zheng, Yiming Xu, Huixiang Liu, Bin Shi, Jiaxiang Wang, Bo Dong, A Survey of Tax Risk Detection Using Data Mining Techniques, DOI: 10.1016/j.eng.2023.07.014,2023
[10] Beatrice Oyinkansola Adelakun, Ebere Ruth Onwubuariri, Gbenga Adeniyi Adeniran & Afari Ntiakoh, Enhancing fraud detection in accounting through AI: Techniques and case studies. DOI:10.51594/farj.v6i6.1232, 2023
[11] Lior Rokach and Oded Maimon, Decision Trees. DOI: 10.1007/0-387-25465-X_9,2005
[12] Soukaina Hakkal, Ayoub Ait Lahcen, XGBoost To Enhance Learner Performance Prediction. DOI: 10.1016/j.caeai.2024.100254,2024
[13] Sakshi Indolia, Anil Kumar Goswami, S. P. Mishra b, Pooja Asopa, Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. DOI: 10.1016/j.procs.2018.05.069,2018
[14] Sepp Hochreiter, Jürgen Schmidhuber, Long Short-term Memory. DOI: 10.1162/neco.1997.9.8.1735,1997
[15] Harsh H. Patel, Purvi Prajapati, Study and Analysis of Decision Tree Based Classification Algorithms, year: 2019. DOI:10.26438/ijcse/v6i10.7478.
[16] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, Michael Weyrich, A survey on long short-term memory networks for time series prediction. DOI: 10.1016/j.procir.2021.03.088,2020
[17] Tianqi Chen, Carlos Guestrin, XGBoost: A Scalable Tree Boosting System. DOI: 10.1145/2939672.293978521. Kumar V., & Singh A. Comparative Study on Various Machine Learning Algorithms for fraud detection in Banking. Int J Com App. 2022 This article describes a comparative analysis of numerous ML techniques for the efficiency in fraud in banking transactions.

[18] Zhang, J., & Zhao, Y. (2021). Review of machine learning methods on fraud detection in online transaction systems: A review. Journal of Internet Services and Applications. This is a review regarding machine learning methods particularized with an ability toward detection of fraud in online transactions within the context of the various challenges and future research perspectives in the field.

[19] Kumar V., & Singh A. Comparative Study on Various Machine Learning Algorithms for fraud detection in Banking. Int J Com App. 2022 This article describes a comparative analysis of numerous ML techniques for the efficiency in fraud in banking transactions.

[20] Almeida, F. C., & Santos, J. R. Deep learning approach for fraud detection in financial transactions. Journal of Financial Technology (2020). This work introduced a deep learning approach to detect fraud in financial transactions, bringing forth the advantages of the application of neural networks.

[21] Bai, Y., & Liu, Y. (2019). A novel approach to fraud detection using machine learning and big data analytics. Journal of Big Data. The paper deals with the integration of big data analytics and machine learning to detect fraud and discusses that big data volume and variety significantly help in enhancing the detection rate.

[22] Sahu, A. K., & Kumar, A. (2021). A review of machine learning techniques for detecting financial fraud. Journal of Financial Crime. This review article discusses different machine learning techniques applied to the detection of financial fraud and hence gives a comprehensive overview of their applications and effectiveness.

[23] Zhou, J., & Wang, H. (2022). A review of machine learning approaches to e-commerce fraud detection. Journal of Electronic Commerce Research. The article is devoted to the field of machine learning methods especially for the task of detecting fraud in e-commerce, discussing unique problems and solutions of this extremely dynamic sphere.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 ✆ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com

Scan to save the contact details