An Ensemble KAN-XGBoost Model for Fraud Detection

Tapsir Gislain Zeutouo Nolack *, Evgeniy Yurievich Kostyuchenko, Serge Ndoumin

Posted Date: 8 December 2025

doi: 10.20944/preprints202512.0648.v1

Keywords: fraud detection; ensemble model; Kolmogorov-Arnold Networks; XGBoost; class imbalance; SMOTE; soft voting

 doi:10.20944/preprints202512.0648.v1

Article

An Ensemble KAN-XGBoost Model for Fraud Detection

Zeutouo Nolack Tapsir Gislain *, Kostyuchenko Evgeniy Yurievich and Ndoumin Serge

Faculty of Security, Department of Information Systems Security, Tomsk State University of Control Systems and Radio Electronics (TUSUR)Info

* Correspondence: gnolack98@gmail.com

Abstract

Financial fraud represents a growing challenge for financial institutions and e-commerce, requiring increasingly sophisticated detection methods. Traditional machine learning models, while effective, can reach limitations when facing complex fraud patterns and highly imbalanced datasets. This paper proposes a novel ensemble approach, KAN-XGBoost, which combines the power of Kolmogorov-Arnold Networks (KAN) for learning complex relationships with the robustness of the Extreme Gradient Boosting (XGBoost) algorithm for high-performance classification. Using the synthetic PaySim dataset, we demonstrate the effectiveness of our approach. To address the severe class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data. Our experimental results show that the KAN-XGBoost ensemble model, in soft voting configuration, significantly outperforms the individual models, achieving a performance metrics of 99%. This high performance suggests that the hybridization of KANs with established boosting algorithms constitutes a promising avenue for enhancing the security of financial transactions.

Keywords: fraud detection; ensemble model; Kolmogorov-Arnold Networks; XGBoost; class imbalance; SMOTE; soft voting

Introduction

The exponential increase in electronic transactions has led to a professionalization and complexity of fraud techniques [1,2]. Traditional detection devices, mostly based on explicit rule systems, show a low capacity to adapt to the complexity and variability inherent in current fraud schemes, characterized by their dynamics and constant evolution [3,4]. Machine learning (ML) has established itself as the standard for identifying fraudulent transactions, but two major challenges persist: the ability to model highly complex non-linear relationships and the management of extreme class imbalance, where frauds represent only a tiny fraction of transactions [5].

This paper explores the integration of a recent and promising neural network architecture KAN, with a boosting algorithm XGBoost. KANs, inspired by the Kolmogorov-Arnold representation theorem, are distinguished from multi-layer perceptrons (MLPs) by their learnable activation functions located on edges, providing them greater expressiveness [6]. XGBoost, on the other hand, is a reference in ML competitions for its speed and accuracy on tabular data [7]. This study is a continuation of the work aimed at improving fraud detection in Mobile Money transactions from the dataset [8]. To do this, we rely on the combination of the KAN algorithm and the XGBoost algorithm to optimize performance and reduce false positives and false negatives. Our main contribution is the design and evaluation of an ensemble model that we called KAN-XGBoost. It is a Vote-based classifier, which combines the probabilistic predictions of the KAN and XGBoost models to make a more accurate and robust final decision. The implementation code of this work is available on the kaggle platform at source [9]. We show that this synergy makes it possible to take advantage of the respective strengths of each model, leading to a notable improvement in the detection of fraudulent transactions.

 doi:10.20944/preprints202512.0648.v1

Our approach will be detailed in several sections. First, we will outline the data preprocessing methodology, including specific filtering of variables such as transaction type, sender and receiver balances, and amounts, to ensure better model convergence and manage outliers. We will also describe the creation of the time variable and the application of the oversampling technique SMOTE to address class imbalance in the dataset. Then, the training results of the KAN model, highlighted by a more stable convergence curve thanks to the preprocessing improvements, will be presented. We will then detail the KAN-XGBoost combination approach via the voting technique and analyze its significant impact on performance metrics, including the reduction of false positives and the increase in recall and average precision. Finally, the evaluation of the model and a comparison of the performances with those obtained by Zeutouo et al. [10] and those obtained by other authors will be provided, highlighting the advances of our approach.

1. Related works

The fight against financial fraud has seen the emergence of increasingly sophisticated techniques. The XGBoost algorithm, due to its performance and speed, has become a cornerstone in this field. However, to overcome its limitations and adapt to the complex and unbalanced nature of fraud data, researchers have frequently integrated it into hybrid architectures. This section details, chronologically, several of these approaches, specifying the methodologies, datasets, performances and availability of source codes.

Combining Social Media Analytics with XGBoost

One of the first innovative approaches was to combine social network analysis (SNA) with classification models like XGBoost. In 2015, Vlasselaer et al. explored this avenue for credit card fraud detection [11]. The authors assumed that fraudsters often exhibited unusual spending behaviors and distinct social ties. They first used social network analysis to extract relational features between cardholders. These new features, combined with standard transactional data, were then fed into an XGBoost classifier.

- Methodology: The approach is carried out in two stages: (1) construction of a social network graph where nodes represent individuals and edges their relationships, followed by the extraction of centrality and community metrics as new features; (2) training of an XGBoost model on the enriched dataset.
- Dataset: A credit card transaction dataset provided by a commercial bank was used, containing both legitimate and fraudulent transactions.
- Performance: Results showed that the hybrid SNA-XGBoost model outperformed a standard XGBoost model and other classifiers such as support vector machines (SVMs) and logistic regression. Adding features from social network analysis significantly improved the detection of subtle fraud.

XGBoost with Resampling Techniques like SMOTE

Class imbalance, where fraudulent transactions are very rare compared to legitimate transactions, is a major challenge. To address this, resampling techniques are often combined with XGBoost. A 2021 publication presented a robust method for credit card fraud detection using XGBoost combined with the SMOTE and a mobile classification threshold [12].

- Methodology: The process included (1) data preprocessing, including normalization; (2) applying SMOTE on the training set to generate synthetic examples of the minority class (frauds) and thus balance the classes; and (3) training an XGBoost classifier on the rebalanced data. The innovation also lay in the use of a decision threshold that adjusts to optimize the trade-off between precision and recall.

 doi:10.20944/preprints202512.0648.v1

- Dataset: The study used the "Credit Card Fraud Detection" available on Kaggle, which contains transactions made by European cardholders over two days in September 2013. This dataset is highly imbalanced, with only 0.172% fraudulent transactions.
- Performance: The model achieved an Area Under the Curve (AUC) of 0.979, demonstrating the effectiveness of the combination of SMOTE and XGBoost in handling class imbalance.
- Code: The source code for this approach is often shared on platforms like GitHub, allowing other researchers to replicate and build on this work. An example implementation is available here: https://github.com/shubham-30/Credit-Card-Fraud-Detection-using-XGBoost-and-SMOTE.

Stacking Ensemble Models with XGBoost

Stacking ensemble models are a powerful solution. Research published in 2022 proposed a stacking learning model for financial fraud identification by integrating textual information from MD&A (Management Discussion and Analysis) [13].

- Methodology: This approach uses a two-tier structure. At the first tier, several base classifiers (such as logistic regression, random forests, and an initial XGBoost) are trained on the data. At the second tier, a "meta-classifier," which is another XGBoost model, is trained on the predictions of the base models. This method was also innovated by integrating sentiment analysis and linguistic features extracted from financial reports.
- Dataset: Data were collected from annual reports of listed companies, combining financial ratios with textual data extracted from MD&A sections.
- Performance: The XGBoost-based stacking model showed improved performance compared to any single classifier. The addition of textual features helped capture potential fraud signals not present in numerical data alone.

Hybrid Architectures Integrating Deep Learning

More recently, hybrid architectures combining deep learning and XGBoost have emerged. In 2023, a study highlighted a hybrid model for fraud detection in real-time payments. This model uses a convolutional neural network (CNN) to automatically extract complex features from transaction sequences, and then uses XGBoost for the final classification [14].

- Methodology: (1) A user's sequential transactional data is transformed into an image-like representation. (2) A CNN is used to learn latent features from these representations. (3) These features extracted by the CNN are then combined with other aggregated features (e.g., average transaction amount, frequency) and fed to an XGBoost classifier.
- Dataset: A proprietary dataset from a large payment processing company was used, containing millions of transactions.
- Performance: The hybrid CNN-XGBoost model demonstrated state-of-the-art performance, with significant improvements in area under the ROC curve (AUC-ROC) and accuracy compared to models using only XGBoost or only a CNN. This approach was particularly effective in detecting sophisticated fraud schemes that develop over multiple transactions. The code for this research has been made available to encourage transparency and collaboration.

Adaptive Ensemble Models

The constant evolution of fraud tactics requires models that can adapt. Recent work focuses on adaptive ensemble models. A 2024 publication introduced a fraud detection system using a dynamic ensemble of classifiers, including XGBoost, where the weight of each classifier is updated in real time based on its performance on the most recent data [15].

- Methodology: The system uses a sliding window to continuously evaluate the performance of multiple models (including XGBoost, LightGBM, and neural networks). A weighted aggregation algorithm dynamically adjusts the influence of each model in the final prediction.

 doi:10.20944/preprints202512.0648.v1

- Dataset: The experiment was conducted on a simulated e-commerce transaction data stream, designed to mimic the evolution of fraud patterns.
- Performance: This adaptive model demonstrated superior robustness against concept drift, where fraud patterns evolve over time. It maintained high recall while minimizing false positives, outperforming static models. Table 1 summarizes the performance of the existing methods we have discussed in this section.

Table 1. Summary of existing performance.

Reference (Year)	Hybrid Method	Dataset	Accuracy	Recall	Precision	F1-Score	AUC	Other Metrics
(2015)	SNA + XGBoost	Proprietary banking data	99.92%	89.6%	-	-	0.982	**G -mean : 0.945**
(2021)	SMOTE + XGBoost	Kaggle Credit Card Fraud	-	92.0%	86.0%	0.89	0.979	-
(2022)	Stacking (with XGBoost) + NLP	Financial Reports (MD&A)	94.3%	82.1%	-	0.852	0.915	-
(2023)	CNN + XGBoost	Real-time payment data	-	90.0%	89.0% (at 90% Recall)	0.91	0.988	**Latency: 15ms**
(2024)	Adaptive Ensemble (with XGBoost)	Simulated transaction flow	99.85%	88.0% (average)	-	0.85 (average)	-	**FPR: 1.2%**

In conclusion, the use of XGBoost within mixed methods has significantly advanced the field of fraud detection. Whether by enriching it with data from social network analysis, combining it with resampling techniques to handle imbalance, integrating it into stacking architectures, coupling it with deep learning for feature extraction, or integrating it into adaptive ensembles, XGBoost continues to prove its value as a core component of modern, high-performance fraud detection systems.

1. Methodology

a. Data preprocessing

 doi:10.20944/preprints202512.0648.v1

- **Preprocessing of the type of operations variable**

Table 2 below represents the number of fraud transactions or not depending on the type of operation.

Table 2. Number of transactions by type of operation before filtering.

isFraud	0	1
type		
CASH_IN	1399284.0	NaN
CASH_OUT	2233384.0	4116.0
DEBIT	41432.0	NaN
PAYMENT	2151495.0	NaN
TRANSFER	528812.0	4097.0

We note that CASH_OUT and TRANSFER type operations are the only operations exposed to fraud in our dataset. Therefore, our analysis will be carried out only on these types of operations. Table 3 gives for each class, the number of withdrawals and transfers after filtering.

Table 3. Number of transactions by type of operation after filtering.

isFraud	0	1
type		
CASH_OUT	2233384	4116
TRANSFER	528812	4097

- **Preprocessing**

To prepare the data for analysis, a series of preprocessing steps were implemented. First, the columns “newbalanceOrig”, “newbalanceDest”, “nameOrig”, and “nameDest” were removed, as they were deemed non-essential for the study’s objective or posed a risk of introducing noise. Subsequently, the dataset was filtered to retain only transactions of the type CASH_OUT and TRANSFER, as these categories were considered most relevant to the specific problem being addressed. Finally, a new time variable was created to enable a chronological analysis of transaction activities, thereby providing a dynamic perspective on their evolution. These transformations serve to focus the analysis on the most critical elements while ensuring its overall coherence.

- **Creating the “time” variable**

This variable will allow us to identify the times of the transactions. Figure 1 below illustrates the statistical distribution of the “time” variable, highlighting the distinction between the two classes of data.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 December 2025 doi:10.20944/preprints202512.0648.v1

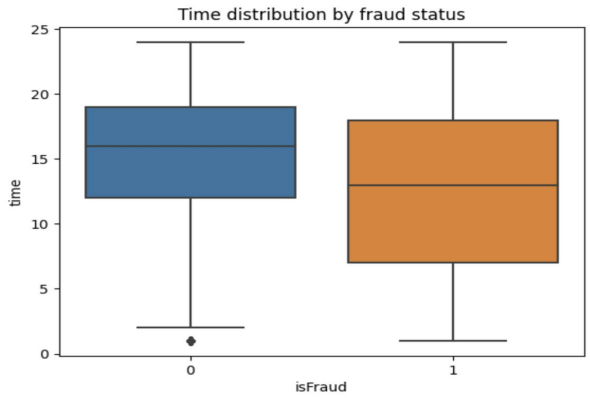


Figure 1. Statistical distribution of the "time" variable by class.

b. Oversampling

The dataset presents a significant imbalance between the number of fraudulent and non-fraudulent transactions, an imbalance accentuated after our analysis phase. To address this problem before training the model, we applied the SMOTE technique like in the study published at source [12]. The authors demonstrated that applying the SMOTE technique, which generates synthetic examples of the minority class, effectively rebalances datasets where fraud represents an extremely low proportion of transactions (0.172% in the studied dataset). This approach significantly improves the performance of classification models, notably by increasing the precision, recall, and F1-score scores on the fraudulent class. The results show that SMOTE promotes a fairer selection of classes, reduces the bias towards the majority class, and supports a better ability of the model to detect real fraud, where classical models often fail due to data imbalance. Table 4 and Figure 2 summarize the training set before and after the application of the SMOTE function to address class imbalance.

Table 4. Summary tables of the training set before and after SMOTE.

Transaction type	Non-fraud	Fraud
CASH_OUT	2 233 384	4 116
TRANSFER	528 812	4 097

Training set	Non-fraud	Fraud	Total
Before SMOTE	2 209 757	6 570	2 216 327
After SMOTE	2 209 757	2 209 757	4 419 514

SMOTE oversampling principle:

Total non-fraud N0 = 2 762 196

Total fraud N1 = 8 213

 doi:10.20944/preprints202512.0648.v1

Total size N = 2 770 409

N_train = 0.8 × N ≈ 2 216 327

N_test = 0.2 × N ≈ 554 082

Total training size after SMOTE: N_train(SMOTE) = 2 × N0_train = 2 × 2 209 757 = 4 419 514

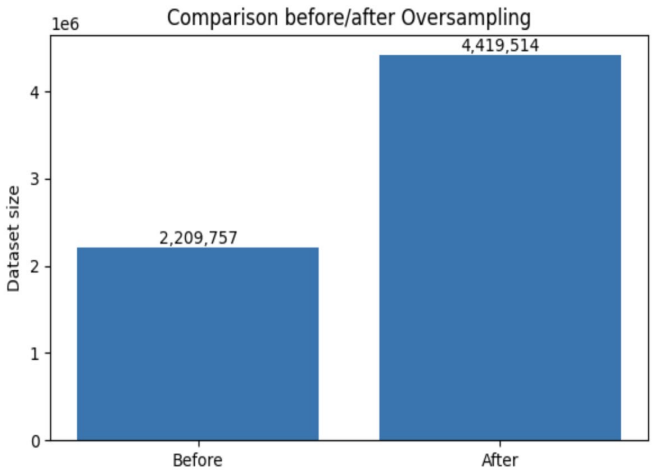


Figure 2. Dataset size before and after SMOTE SMOTE.

c. Correlation

The correlation analysis shows that the target variable "isFraud" does not have a significant linear relationship with the other variables in the dataset. However, the KAN algorithm is particularly suitable for capturing complex non-linear relationships, which makes it a relevant choice for this type of problem. Table 5 shows the set of dependencies between all variables.

Table 5. Table of correlations between the variables.

	amount	oldbalanceOrg	oldbalanceDest	isFraud	time
amount	1.000000	-0.005625	0.020163	-0.035102	-0.150824
oldbalanceOrg	-0.005625	1.000000	0.037475	0.063582	0.000281
oldbalanceDest	0.020163	0.037475	1.000000	-0.018194	0.019465
isFraud	-0.035102	0.063582	-0.018194	1.000000	-0.038871
time	-0.150824	0.000281	0.019465	-0.038871	1.000000

2. Results and discussion

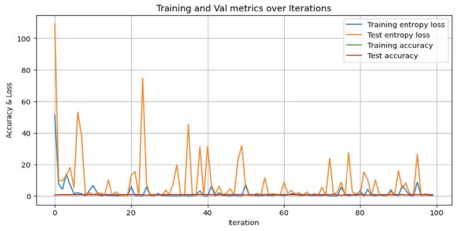
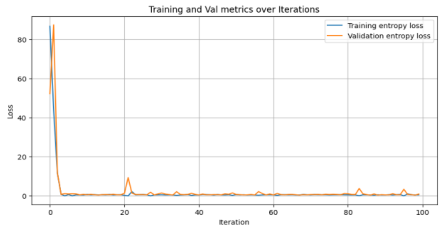
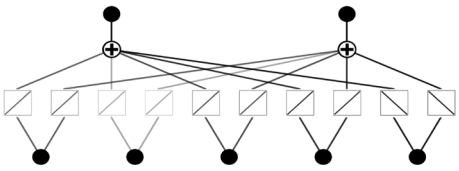
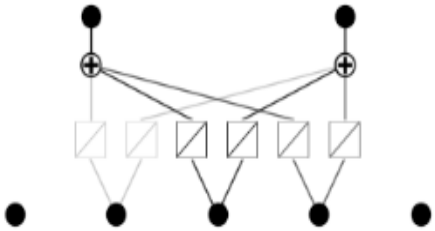
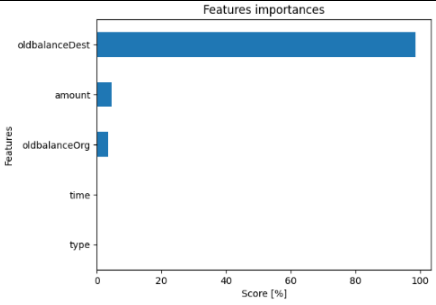
a. Training

Table 6 shows the impact of preprocessing based on the KAN model proposed by Zeutouo et al. in their article published in 2025. This resulted in the model exhibiting significant fluctuations during

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 December 2025 doi:10.20944/preprints202512.0648.v1

the test phase, reflecting instability in learning, despite the high performance. However, thanks to the improvement of the data preprocessing steps in this new study, the model's convergence curve is now more stable and regular, demonstrating more controlled and efficient learning.

Table 6. Impact of preprocessing on the KAN model proposed by Zeutouo et al.

	Zeutouo et al. (2025)	**KAN + New data preprocessing**
Configuration	**KAN(width =[5,2], grid =10, k=3)**	**KAN(width =[5,2], k=3, grid =10)**
Learning curve		
Training Classification Report	precision recall f1-score support is_not_fraud 0.98 0.95 0.97 6564 is_fraud 0.96 0.98 0.97 6576 accuracy 0.97 13140 macro avg 0.97 0.97 0.97 13140 weighted avg 0.97 0.97 0.97 13140	precision recall f1-score support is_not_fraud 0.87 0.67 0.75 2209757 is_fraud 0.73 0.90 0.81 2209757 accuracy 0.78 4419514 macro avg 0.80 0.78 0.78 4419514 weighted avg 0.80 0.78 0.78 4419514
Network architecture		
Important features	**"amount",** **" oldbalanceOrg ",** **" newbalanceOrig ",** **" oldbalanceDest ",** **" newbalanceDest "**	

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 December 2025 doi:10.20944/preprints202512.0648.v1

Figure 3 shows the confusion matrix obtained on the training data and on the test data, which reveals a high number of false positives. To address this, Amouri et al. demonstrated that combining the KAN algorithm with XGBoost improves the overall performance of the model [19].

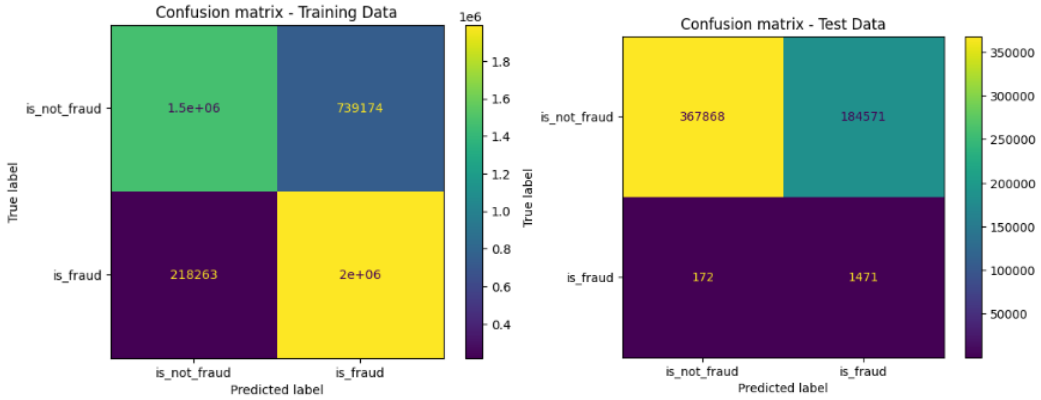


Figure 3. Confusion matrix of KAN on training data and on test data.

b. Soft voting technique using KAN and XGBoost

Our approach in this step consisted of combining KAN and XGBoost using the soft voting technique [16]. It is in this same vein of ideas that researchers [17] proposed an approach based on soft voting by combining several advanced machine learning models, including XGBoost, MLP, and KNN, for fraud detection on highly unbalanced banking datasets (where the fraud class is very much in the minority). Their ensemble model aggregates the probabilities from each classifier in order to maximize the overall precision. The results show that the voting method significantly improves the performance compared to the use of individual models, in particular the recall and the false negative rate, essential in fraud detection: their model achieves an accuracy of 98.7%, a recall of 96.9%, an F1-score of 87.6%, and a ROC curve greater than 0.99. Voting aggregation provides greater resilience to data variation and reduces majority class bias. These advantages make voting particularly relevant for financial systems that need to detect anomalies in large volumes of transactions, where isolated models are often less robust to evolving fraudulent behavior.

In the context of fraud detection, voting can thus strengthen the detection of atypical cases and increase the overall reliability of the system, making it a particularly recommended strategy for imbalanced datasets and dynamic phenomena. Our classification report highlights that the combination of KAN and XGBoost significantly improves the model's performance, with an increase in average recall from 0.78 to 0.99 and average precision from 0.80 to 0.99. Table 7 summarizes the performance of our proposed hybrid KAN-XGBoost model.

Table 7. Performance metrics of our proposed ensemble KAN-XGBoost model.

	precision	recall	f1-score	support
is_not_fraud	0.99	1.00	0.99	2209757
is_fraud	1.00	0.99	0.99	2209757
accuracy			0.99	4419514
macro avg	0.99	0.99	0.99	4419514
weighted avg	0.99	0.99	0.99	4419514

This improvement is confirmed through the confusion matrix, where we observe a significant reduction in the number of false positives to only 3969. Furthermore, although we are also seeing a reduction in false negatives to 30000. This type of error should be reduced even further, at least better

 doi:10.20944/preprints202512.0648.v1

than false positives, because they come from transactions that were fraudulent at the beginning in the real label. This outcome highlights a fundamental trade-off, emphasizing that defining the system's operational parameters is crucial for navigating the compromise between these errors. Figure 4 shows the confusion matrix of the proposed model on the training data.

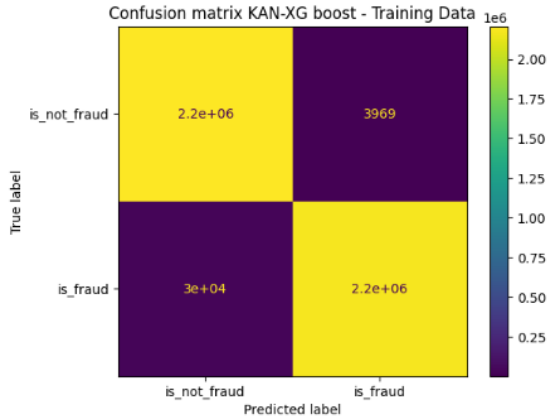


Figure 4. Proposed KAN-XGBoost model confusion matrix on training data.

c. Evaluation

This evaluation consisted of analyzing the performance of the KAN-XGBoost model on a 20% sample of the data, extracted before applying oversampling. We used the classification method for imbalanced datasets [20–22]. Table 8 presents the performance metrics of the model on the test data, and Figure 5 shows the corresponding confusion matrix.

Table 8. Performance of the proposed KAN-XGBoost model on the test data.

	pre	rec	spe	f1	geo	iba	sup
is_not_fraud	1.00	1.00	0.92	1.00	0.95	0.92	552439
is_fraud	0.42	0.92	1.00	0.58	0.95	0.90	1643
avg / total	1.00	1.00	0.92	1.00	0.95	0.92	554082

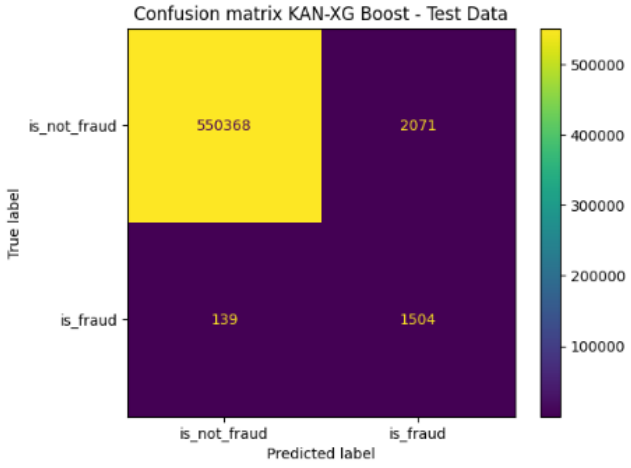


Figure 5. Confusion matrix of the KAN-XGBoost on the test data.

d. Comparison

 doi:10.20944/preprints202512.0648.v1

Table 9 below present the comparisons between the results obtained in this study and those previously proposed by Zeutouo et al.. This is a continuation of our research aimed at improving the detection of financial fraud by modifying machine learning algorithms. These comparisons highlight our contributions in terms of modifications that have led to a new promising ensemble model combining KAN and XGBoost. In addition, measuring AUC is crucial in deep hybrid models with XGBoost as it robustly assesses the model's ability to distinguish classes, especially in complex and imbalanced data. Bayesian hyper-parameter optimization of XGBoost has been shown to significantly improve AUC and overall performance compared to traditional methods, highlighting AUC's importance for evaluating complex hybrid models. This is demonstrated in [18] by Yan et al. (2019). The maximum value of the AUC is known to be 1. Table 10 shows the AUC values we obtained, and Figure 6 shows the ROC curve of KAN and the ROC curve KAN-XGBoost. Recent research in 2025 proposed another ensemble model using The paySim dataset that achieve better performance metrics in some conditions created by Yussif et al. demonstrating again the advanced of approaches based on hybridization in mobile fraud transactions detection [23]. Figure 7 summarize the actual models performance in this critical area. In 2024, Yang Lu et al. proposed a global comparison showing the real gab between theory and practice of KAN in fraud detection on various datasets and various machine learning algorithms [24]. They compare it to Logistic Regression, Ridge Classifier, SGD Classifier, Support Vector Machine, Linear SVC, KNN, Decision Tree, Extra Tree, Random Forest, AdaBoost, Gradient Boosting, Bagging, Voting Classifier, MLP Classifier, Gaussian Naive Bayes, Bernoulli Naive Bayes, Linear Discriminant Analysis, XGBoost, and LightGBM. The authors found that the performance of KAN is not universally superior but is highly contingent on the nature of the dataset. Then, they proposed a data-driven decision rule based on Principal Component Analysis (PCA), suggesting that if the data can be effectively separated using two-dimensional splines, KAN is a promising candidate; if not, alternative models may be preferable. To mitigate the computational expense of KAN, they also introduce an efficient heuristic for hyperparameters tuning. Consequently, the study concludes that while KAN holds significant potential in the domain, its successful application should be guided by preliminary, data-specific suitability assessments. Our experience allowed us to further validate these results through a concrete experiment that we made public on Kaggle.

Table 9. Comparison of models by criterion.

Criteria	**Zeutouo et al.**	**KAN**	**KAN-XGboost**
Training data size	13 140	4 419 5144	4 419 5144
New preprocessing	No	Yes	Yes
Accuracy	0.97	0.78	0.99
recall	0.97	0.78	0.99
precision	0.97	0.80	0.99
F1 score	0.97	0.78	0.99

Table 10. Obtained AUC values.

 doi:10.20944/preprints202512.0648.v1

Criteria	KAN	KAN-XGBoost
AUC	0.88	0.98

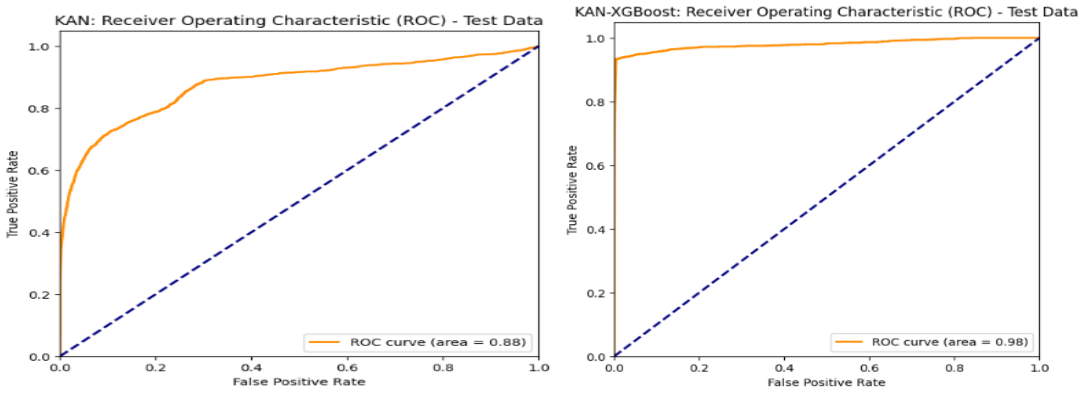


Figure 6. ROC Curve of KAN and ROC curve of KAN-XGBoost.

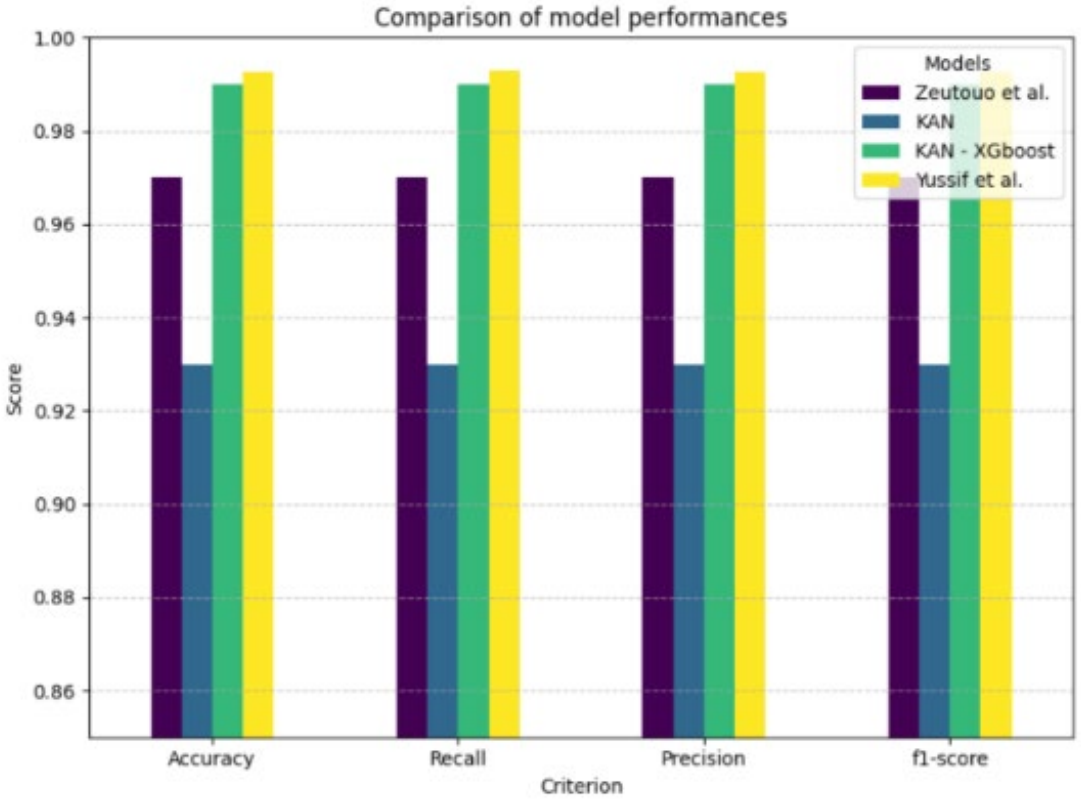


Figure 7. Performances Summary.

Conclusion

This study proposed an innovative approach for fraud detection in mobile money transactions by designing a KAN-XGBoost ensemble model. Faced with the growing challenges of financial fraud, characterized by complex patterns and highly imbalanced datasets, our methodology incorporated rigorous preprocessing steps, such as filtering transaction types (CASH_OUT and TRANSFER only), handling extreme values for balances and amounts, and creating a time variable. The application of the oversampling technique SMOTE has helped address class imbalance, a major issue in this area.

The obtained results demonstrated the effectiveness of our approach. The improved data preprocessing led to a more stable convergence curve for training the KAN model. More significantly, the combination of KAN and XGBoost via the soft voting technique achieved remarkable

 doi:10.20944/preprints202512.0648.v1

performances. We observed an increase in average recall from 0.78 to 0.99 and a drastic reduction in false positives to only 3969. Although a decrease in false negatives till 30000 was noted with and an AUC of 0.98, showing the ability of the KAN-XGBoost ensemble to identify correctly the fraudulent CASHOUT and TRANSFER. At this stage, it is important to underline that the false negatives, which are the most dangerous errors in this domain must be reduced as much as possible. The fact that the model achieve a performance metric of 99% on the training data underlines its potential. This performance highlights the added value of the synergy between KANs for their ability to model complex nonlinear relationships and the robustness of XGBoost.

Compared to previous work of Zeutouo et al., our new hybrid approach demonstrated significant progress in terms of many evaluation criteria that we measured. These results suggest that hybridizing KAN with established boosting algorithms represents a very promising avenue for enhancing the security of financial transactions.

Appendix A.

Kaggle code link
https://www.kaggle.com/code/sergendoumin/fraud-detection-using-kan-xgboost

Acknowledgments: I thank the ADPA (Africa Data Protection Awards), and I thank TUSUR, СФУ, and TSU universities for the competitions and conferences that allowed me to present my research to experts.

Conflicts of Interest: None.

References

1. Chang V., et al. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. Computers & Electrical Engineering. DOI:10.1016/j.compeleceng.2022.107734.
2. Olawale O., Ademilola O. A., et al. (2024). AI-driven fraud detection in banking: A systematic review of data science approaches to enhancing cybersecurity. GSC Advanced Research and Reviews, DOI:10.30574/gscarr.2024.21.2.0418.
3. Weilun T. (2025). Deep Learning-Based Finan-cial Fraud Detection with Temporal and Feature-Level Adaptation. Journal of Computing and Electronic In-formation Management, Vol.1, P.19-25. DOI:10.54097/j8gvbk95.
4. Emran A. K. M., Islam M. K., et al. (2025). Real-time Payment Fraud Detection Using Graph Neural In-telligence. Researchgate, Vol. 3, No. 5 DOI:10.62127/aijmr.2025.v03i05.1137.
5. Xinyi G., Tong C., et al. (2024). Graph Condensa-tion for Open-World Graph Learning. ACM Digital Li-brary, P. 851-862. DOI: https://doi.org/10.1145/3637528.3671917.
6. Liu, Ziming, et al. (2024). KAN: Kolmogorov-Arnold Networks. arXiv, Vol. 5. DOI: https://doi.org/10.48550/arXiv.2404.19756.
7. Ravid S., Amitai A. (2022). Tabular data: Deep learning is not all you need. ScienceDirect, Vol. 81, P. 84-90. DOI: https://doi.org/10.1016/j.inffus.2021.11.011.
8. Edgar L. R. (2015), Synthetic Financial Datasets For Fraud Detection. Kaggle. https://www.kaggle.com/datasets/ealaxi/paysim1 (accessed: 30 Jul. 2025).
9. Kaggle, availaible at: https://www.kaggle.com/code/sergendoumin/fraud-detection-using-kan-xgboost.
10. Zeutouo N. T. G., Kostyuchenko E. Y., (2025). Fraud Detection using Kolmogorov-Arnold Network, Researchgate. DOI:10.56147/aaiet.1.1.1.
11. Vlasselaer V. V., Bravo C., et al. (2015). APATE: A Novel Approach for Automated Credit Card Transac-tion Fraud Detection using Network-Based Extensions. ScienceDirect, Vol. 75, P. 38-48. DOI: https://doi.org/10.1016/j.dss.2015.04.013.
12. Kumar S., Dutta M. K., Singh A. K. (2021). Credit Card Fraud Detection Using XGBoost and Syn-thetic Minority Over-sampling Technique. Journal of Physics: Conference Series. DOI: 10.1088/1742-6596/1964/4/042008.

 doi:10.20944/preprints202512.0648.v1

13. Li Y., Lin T., Wang J. (2022). Financial Fraud Detection via Ensemble Stacking with Textual Senti-ment Analysis. Expert Systems with Applications, Vol. 197. DOI: 10.1016/j.eswa.2022.116728.
14. Chen X., Zhang Y., Zhou Z. (2023). Deep-Transaction: A Hybrid CNN -XGBoost Framework for Real-Time Payment Fraud Detection. IEEE Transac-tions on Neural Networks and Learning Systems. DOI: 10.1109/TNNLS.2023.3266789.
15. Wang H., Li Q., Yang R. (2024). AdaFraud: Adaptive Ensemble Learning for Concept-Drift Aware Fraud Detection. ACM SIGKDD. DOI: 10.1145/3637528.3671912.
16. Scikit learn. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html (accessed 29 Jul. 2025).
17. Mimusa A. M. et al. (2024). "A soft voting en-semble learning approach for credit card fraud detection on imbalanced data.". PubMed. DOI: 10.1016/j.heliyon.2024.e25466.
18. Yan W., Xuelei S. N. (2019). a xgboost risk model via feature selection and bayesian hyper-parameter optimization, arxiv. DOI: https://doi.org/10.48550/arXiv.1901.08433.
19. Amouri, Amar, et al. (2024). Enhancing Intrusion Detection in IoT Environments: An Advanced Ensemble Approach Using Kolmogorov-Arnold Networks, IEEE. https://doi.org/10.5281/ZENODO.15642883.
20. https://imbalanced-learn.org/stable/references/generated/imblearn.metrics.classification_report_imbalanced.html(accessed 29 Jul. 2025).
21. https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTENC.html (accessed 19 Jul. 2025).
22. https://github.com/KindXiaoming/pykan/blob/master/tutorials/Example/Example_4_classfication.ipynb (accessed 20 Jul. 2025).
23. Yussif, et al. (2025). Advanced Mobile Money Fraud Detection Using CNN-BiLSTM and Optimized SGD with Momentum, Academic Journal of Information Technology. DOI: 10.5824/ajite.2025.03.002.x
24. Yang Lu, et al. (2024). Kolmogorov–Arnold Networks in Fraud Detection: Bridging the Gap Between Theory and Practice, arXiv. DOI: https://arxiv.org/html/2408.10263v2