

Review of Computer Engineering Research

2026 Vol. 13, No. 4, pp. 100-118

ISSN(e): 2410-9142

ISSN(p): 2412-4281

DOI: 10.18488/76.v13i4.5231

© 2026 Conscentia Beam. All Rights Reserved.



Privacy-preserving hybrid AI framework for secure fraud detection in .NET applications

 Venkatesh
Ramasamy^{1*}

 Saranya Devi
Annadurai²

 Anusuya
Venkatasamy³

 Rajesh Kumar
Vikramathithan⁴

¹School of Computing, SRM Institute of Science and Technology
Tiruchirappalli, Tiruchirappalli, Tamil Nadu, India.

²Email: rvenkateshassistantprofessor@gmail.com

³Department of Computer Science and Engineering (Cyber Security), Sethu
Institute of Technology, Pulloor, Kariapatti, Virudhunagar, 625115, India.

⁴Email: saranyadevi1188@gmail.com

⁵Department of Computer Science and Engineering, Ramco Institute of
Technology, North Venganallur Village, Rajapalayam 626117, India.

⁶Email: pgkrishanu@gmail.com

⁷Department of Electrical and Electronics Engineering, Sir M. Visvesvaraya
Institute of Technology, Bengaluru, 562157, India.

⁸Email: rajeshkumar_eee@sirmvit.edu



(+ Corresponding author)

ABSTRACT

Article History

Received: 3 April 2026

Revised: 22 June 2026

Accepted: 13 July 2026

Published: 7 October 2026

Keywords

ASP.NET integration
Credit card fraud detection
Deep learning
Hybrid ensemble model
PAHEM
Privacy-preserving machine
learning
Secure data analytics
XGBoost.

The fast growth of digital payment systems has made the problem of credit card fraud a major threat that demands a smart and secure system for detecting fraud. Certain cases of machine learning are known to be very bad at identifying fraud with severe imbalance in classes and complex behavior patterns in financial data. This paper suggests a Privacy-Aware Hybrid Ensemble Model (PAHEM) on a secure.NET web application framework for real-time fraud analytics. The proposed strategy involves the use of the Random Forest-based feature ranking, Extreme Gradient Boosting (XGBoost), and a Deep Neural Network (DNN) to identify both the linear and nonlinear trends in the transaction data. In order to eliminate the imbalance in classes, the Synthetic Minority Oversampling Technique (SMOTE) is used when training a model. The trained hybrid model is deployed as an encrypted analytics service that may be attached to the ASP.NET-based financial systems through the resources of the REST API, which will help to identify the fraud risk in real-time when processing transactions. Experimental analysis (on publicly available credit card fraud data) shows that the proposed framework can achieve a high detection accuracy of 99.94, 0.8229 precision, 0.8061 recall, 0.8144 F1-score, 0.9997 specificity, and ROC-AUC of 0.9819. These findings suggest that the offered PAHEM model suggests a high-quality detection of fraud and a small number of false positives. The combination of hybrid AI models with secure.NET web applications provides a scalable and privacy-conscious architecture that can be implemented in real-life financial transaction monitoring systems.

Contribution/Originality: This research is original by combining a hybrid ensemble (Random Forest, XGBoost, and DNN) with SMOTE for handling imbalanced data in a privacy-preserving and encrypted manner. Unlike offline model evaluation, a .NET REST API for real-time credit card fraud detection allows for secure and scalable deployment within financial systems.

1. INTRODUCTION

The rate of development of online payment platforms and other digital financial services has drastically augmented the number of electronic transactions globally. Although these technologies are convenient and efficient, they also create severe security issues, especially in the shape of credit card fraud and financial cybercrime. Financial transactions lead to fraud cases that cause significant economic losses to financial institutions and customers annually.

As recent reports suggested, the sophistication and number of fraud attacks keep increasing because of the development of automated attack methods and the extensive use of digital payment infrastructures. Conventional rule-based fraud detection systems may not be able to overcome these problems since they are usually based on fixed patterns and are not able to adjust to new fraud patterns. This has seen machine learning and artificial intelligence methods attract significant interest in creating intelligent fraud detection machines that can detect complex transactions and anomalies. Nonetheless, to use such models efficiently, it is necessary to address further difficulties like data privacy, the type of classes, and the compatibility of these systems with the secure web-based systems of financial operations. The high rate of online network financial systems, cloud computing environments, and online service infrastructures has greatly augmented the amount of sensitive data being handled in dispersed settings. Although these technologies allow for significant advantages in terms of scalability, efficiency, and accessibility, they also present significant security and privacy issues. Specifically, the increased use of machine learning and artificial intelligence systems to conduct data analytics has created concerns about the security of sensitive data when training models, performing inference, and deploying them. The recent research has highlighted the need to create privacy-aware data analytics systems, which are capable of ensuring data confidentiality and also permit proper and scalable intelligent decision-making. As an example, Saini et al. [1] championed a hybrid machine learning system composed of an optimization algorithm to increase trust and privacy protection in the cloud setting, and evidenced the significance of incorporating the best analytics with safe computing systems.

As more AI systems find applications in healthcare and other sectors of the economy, such as finance and industry, privacy-preserving artificial intelligence has become a research priority. Khalid et al. [2] introduced an extensive review of the privacy-sensitive AI tools developed in the healthcare system and demonstrated the necessity of secure machine learning frameworks that can ensure the protection of sensitive personal data. In the same way, Kaissis et al. [3] studied the concept of secure and federated learning in medical imaging using a distributed learning strategy by highlighting how such methods can produce collaborative model training without exposing raw data. These methods have demonstrated to be useful in securing sensitive datasets and still achieve a high predictive rate. Nevertheless, the deployment of these privacy-protective measures in actual data analytics systems is not an easy affair because of the complexity of calculations, scalability problems, and integration options in operational software systems. In order to curb the issue, scholars have examined some of the methods of ensuring the privacy of data in machine learning systems. El Mestari et al. [4] have addressed the different measures of maintaining privacy of data in the machine learning process, such as encryption schemes, secure multiparty computation, and privacy-aware model design. Moreover, with the introduction of massive generative AI models, the issue of possible privacy breaches and unauthorized exposure of data has become highly alarming. A detailed survey of privacy and security risks in generative AI systems by Golda et al. [5] highlights the need to develop strong privacy-protecting measures of AI-based analytics. Recent developments have also considered privacy inference on the client. As a typical example, Auer et al. [6] showed an image classification method with deep learning that can be used in privacy protection and keep sensitive data on the devices of users for direct processing. Privacy-preserving analytics have gained special relevance in financial and fintech applications because the information about a transaction is especially sensitive and regulatory bodies mandate the protection of data. Ju [7] introduced a risk prediction and sensitive information detection privacy preservation framework that was a hybrid model in its approach to fintech and emphasized the usefulness of integrating machine learning strategies with secure analytics procedures. On the same note, Kalpana and Rao [8] have created a hybrid deep learning system that integrates swarm optimization algorithms to facilitate safe and effective retrieval of data, as well as maintenance of the privacy of data, in the cloud. Despite these studies providing encouraging advances in privacy-aware machine learning, the vast majority of the current methods are mainly dedicated to the enhancement of algorithms and seldom consider the practical implementation of those versions into the real web applications setting.

Contemporary enterprise software is relying more on web-based application platforms like ASP.NET and .NET Core to support mass data processing, transacting, and securely offering services. The combination of privacy-conserving artificial intelligence models and such web-based infrastructures is also a challenge to the implementation of intelligent analytics systems in real-life scenarios. Specifically, the secure communication, the data anonymization, and the effective inference of the model will have to be properly implemented in such a way that not only can the performance of the system be guaranteed, but also the privacy. This research will solve these problems by introducing the Privacy-Aware Hybrid Ensemble Model (PAHEM) that will be combined with an architecture of a secure .NET web application to provide privacy-aware data analytics. The proposed solution is a combination of random forest-based feature ranking, Extreme Gradient Boosting (XGBoost), and a Deep Neural Network (DNN) to obtain intricate nonlinear data patterns in transactional data and powerfully classify the data. Besides that, the framework is supplied with privacy-sensitive analytics functions and secure API-based communication that allows an easy integration with ASP.NET web applications. The hybrid model is supposed to enhance the accuracy of fraud detection, scale the system, and minimize privacy risks when conducting data analytics. This study adds to the existing body of literature by introducing a privacy-aware hybrid ensemble approach to fraud detection in financial transactions. This study has a new combination of Random Forest feature selection, XGBoost, and Deep Neural Networks in a secure .NET architecture. This research is one of the few studies that examine the real-time implementation of privacy-preserving AI models in web-based applications. The paper is the first to provide a logical integration of hybrid AI models with ASP.NET secure analytics systems. The key contribution of this paper is that hybrid ensemble learning enhances accuracy in fraud detection while maintaining privacy. This paper provides evidence of the success of the integration of machine learning and secure system deployment in financial analytics.

The key contributions of this work are as follows:

One such system is a privacy-conscious hybrid ensemble learning (PAHEM), which is a mixture of tree-based and deep learning models that can be used to effectively detect fraud.

A safe analytics architecture, which was built on ASP.NET web applications, made it possible to deploy it in practical financial systems.

An overall comparison with a variety of performance measures, such as accuracy, precision, recall, specificity, F1-score, and ROC-AUC.

Comparison of privacy-related measurements, including privacy leakage rate, encryption overhead, and training efficiency of secure AI deployment.

In general, the presented framework shows that the reliability, scalability, and security of the contemporary data analytics systems could be highly enhanced with the help of privacy-preserving hybrid AI models and secure .NET web application infrastructures. Although there are considerable advancements in hybrid machine learning models and secure data analytics, current research mainly addresses algorithmic innovation or application-specific domains, such as healthcare or IoT systems. Few works have explored the deployment of privacy-preserving hybrid AI models in secure real-world web application environments, such as financial fraud detection. Additionally, current studies typically neglect implementation aspects like secure API integration, real-time data processing, and scalability.

2. LITERATURE SURVEY

The works of recent researchers have investigated different machine learning and deep learning methods of enhancing the accuracy of fraud detection in financial transaction systems. Logistic Regression, Support Vector Machines, and Random Forest are old classification algorithms that have been widely used to identify fraud trends in transactions because of their interpretability and efficiency. Nevertheless, the models frequently find it difficult to work with highly unbalanced data in which fraudulent transactions are a small percentage of all transactions. To counter this problem, scientists have suggested a more advanced methodology, which includes ensemble learning, hybrid machine learning structures, and deep neural networks to understand complicated nonlinear relationships

among transaction data. Furthermore, privacy-preserving mechanisms such as secure data analytics, federated learning, and differential privacy have also been proposed in various studies to guarantee that the sensitive financial data is not compromised either when training a model or when making inferences. For real-time fraud detection, the challenge of incorporating privacy-conscious machine learning into safe web application systems continues to be a significant research issue in spite of the mentioned advancements. Machine learning that has preserved privacy has emerged as a core research topic because the use of intelligent analytics systems is on the rise in healthcare, financial, industrial internet of things, and cloud. Numerous recent works have aimed at creating hybrid machine learning systems that integrate predictive precision and secure data processing systems. Almaiah et al. [9] have come up with a lightweight hybrid deep learning model that is applicable to the Industrial Internet of Medical Things (IIoMT). Their strategy combines the effects of deep learning and privacy-sensitive algorithms to provide secure data transfer and analytics in the hospital setting. The work revealed that hybrid deep learning models could be used to enhance the performance of prediction without violating the privacy of sensitive medical information. As the complexity of cyber threats in connected systems grows, more advanced AI-based models have been investigated in the area of threat detection in a secure way. Ferrag et al. [10] proposed a privacy-conscious IoT and Industrial IoT network cyber threat detection BERT-based lightweight model. They employed large language models to examine threat patterns with built-in privacy-conscious products to avoid the leakage of sensitive information. The research demonstrated the opportunities of integrating the deep learning structures with the privacy-saving methods to enhance the detection of threats in large network settings of large settings.

Several studies have also delved into the area of statistical learning and optimization-based methods of privacy-preserving data analytics. Alrayes et al. [11] designed a privacy-preservation system for an IoT network by combining statistical learning with optimization algorithms to address a high-dimensional data space. Their solution was found to be more secure in terms of data and effective performance of analytics in large-scale distributed systems. Likewise, Han et al. [12] suggested a privacy-saving digital marketing system, which analyzes user behavior using deep learning models on data-monetization platforms. Their work stressed the need to establish privacy-conscious AI systems that are capable of processing sensitive user data and still achieving analysis precision. The tremendous growth of generative artificial intelligence and large language models has brought new security and privacy issues in AI systems. A survey of security and privacy risks of large language models was performed by Das et al. [13], and such vulnerabilities as data leakage, adversarial attacks, and threats of model inversion were outlined. Their paper highlighted the importance of strong privacy-saving solutions in the implementation of AI models in sensitive fields of application. Kumar et al. [14] came up with a related study where a hybrid meta-heuristic optimization framework for privacy-preserving industrial IoT systems was proposed. Their model combines smart optimization planning with machine learning algorithms to enhance the security and performance of the systems. Besides algorithmic methods, there have been studies on distributed privacy-preserving analytics platforms. Welten et al. [15] presented an analytics platform for healthcare data that can be distributed to be analyzed collaboratively without breaching patient privacy. Their model employs the use of secure distributed computing methods to work on healthcare data without revealing raw sensitive information. This kind of architecture shows that privacy-conscious machine learning is critical when deployed on a distributed system. The topic of hybrid machine learning frameworks for predictive analytics in healthcare IoT settings has also received recent research studies. Ganji et al. [16] suggested an example of a hybrid machine learning model combined with data analytics methods to perform predictive maintenance of healthcare IoT systems with privacy. Their method proves that the hybrid models are able to improve the prediction rates greatly and provide the safety of sensitive data processing. Wanza et al. [17] present a federated learning framework for fraud detection with homomorphic encryption. They then allow multiple financial institutions to collaborate in the training of their models without releasing any sensitive transactional data, while preserving the detection results and providing strong privacy protection. Likewise, Tyagi [18] covers the potential of privacy-enhancing AI in finance, highlighting the balance of usefulness, security, and regulatory adherence. The study points

out that although AI tools can help detect fraud more accurately, privacy and compliance with regulations are still significant hurdles in implementation. With the idea of extending the federated learning paradigm, Ramalingam et al. [19] present a hybrid federated learning approach with generative AI capabilities for smart environments with an IoT component. Their paper shows that in highly distributed systems, synthetic data generation can deliver better model robustness without compromising privacy. Towards this end, Uzoatu et al. [20] suggest a privacy-preserving federated learning approach for the context of detecting fraud in a distributed banking network. The findings demonstrate that they can successfully mitigate data leakage risk while achieving the fraud detection accuracy necessary for competing financial institutions when implementing decentralized learning. In addition to financial systems, Alotaibi [21] proposes an application-driven secure federated learning framework for smart city applications, namely, parking management based on vision. The results of this study demonstrate how versatile and adaptable the hybrid privacy-preserving architectures can be in various fields, underscoring the scalability of federated and hybrid deep learning approaches. Singh et al. [22] then discuss privacy-preserving mechanisms for resilient digital identity management in cyber-physical systems using AI. They focus on incorporating encryption techniques, authentication protocols, and AI models to enhance the security of identity-centric applications. Lastly, Chaturbhuj et al. [23] explore the use of AI and cryptographic methods for creating cyber-physical systems with security and explainability in IoT environments. Their analysis highlights the need to ensure that explainability is integrated with trust-building AI models designed with security in mind for real-world application.

Even though the current research has achieved great advancements in privacy-aware machine learning and hybrid AI solutions, most of them confine themselves to a domain-specific scenario, like healthcare, industrial IoT, or cybersecurity systems. The challenge of integrating privacy-conscious hybrid machine learning models with secure web-based application platforms of financial transaction analytics has been taken up with little attention. Specifically, real-world deployment settings like ASP.NET and web applications based on the .NET platforms demand effective and scalable AI models capable of detecting fraud in real-time without seriously compromising mechanisms of privacy protection. To fill this gap in the research, the current paper suggests a Privacy-Aware Hybrid Ensemble Model (PAHEM) embedded into the secure web application infrastructure. The framework suggested in the paper provides the power of hybrid machine learning and privacy-conscious analytics to identify fraudulent financial transactions based on the publicly available dataset of credit card fraud, offered by the Machine Learning Group of Université Libre de Bruxelles, which is supplied by Kaggle. This data set includes more than 284,000 anonymized credit card transactions in it, and it has become a popular standard in fraud detection studies. The proposed solution seeks to offer a scalable and privacy-preserving solution to modern financial fraud detection systems by integrating the hybrid AI models and a secure application deployment mechanism. In spite of the tremendous advancements in privacy-conscious machine learning and secure data analytics, various limitations exist in the available research. The majority of existing solutions target domain-specific applications, including healthcare systems, industrial IoT networks, cybersecurity frameworks, and privacy-preserving mechanisms, which are implemented primarily on the algorithmic level. Although these works show that hybrid machine learning models and privacy-sensitive frameworks can be effective in ensuring the safety of analytics, they do not provide a practical implementation plan to apply to the actual software environment. Specifically, the majority of existing solutions lack a solution for the incorporation of privacy-aware artificial intelligence models and secure web application systems, which are prevalent in enterprise systems to support financial transactions and massive data services.

In addition, there is limited literature that focuses on studying a unified framework capable of balancing between accuracy in prediction, privacy protection, and deployability in a secure web platform. The second weakness is that the majority of studies fail to assess the privacy-related operation characteristics of privacy leakage risk, encryption overhead, and system integration efficiency, which are essential to implement in financial analytics systems. Consequently, there exists a necessity to have a scalable hybrid AI system capable of delivering high fraud detection rates as well as being able to offer high privacy assurances, and being able to integrate well with the existing modern

web-based application infrastructure. To fill these knowledge gaps, this project suggests the Privacy-Aware Hybrid Ensemble Model (PAHEM) that will be implemented in conjunction with a secure ASP.NET-based web application framework. The suggested structure is the combination of various machine learning algorithms, such as the Random Forest feature selection, Extreme Gradient Boosting (XGBoost), and a Deep Neural Network, aimed at capturing the complex patterns of transactions without violating the privacy of data. Moreover, the framework has in-built secure analytics systems as well as API-based integration, to facilitate effective implementation in the domain of the .NET web applications to represent real-time fraud detection systems. The proposed PAHEM framework has a detailed methodology and architecture, which are described in the next section.

3. PROPOSED METHODOLOGY

3.1. Overview of the Proposed Framework

In order to overcome the obstacles of precise fraud detection and maintain data privacy and allow real-world implementation, this paper suggests the Privacy-Aware Hybrid Ensemble Model (PAHEM) which is combined with a secure analytics framework. The given model is an integration of several machine learning methods that can identify the intricate price patterns of transactions and be resistant to class imbalance and privacy threats. The architecture is a combination of Random Forest-based feature importance ranking, Extreme Gradient Boosting (XGBoost), and a Deep Neural Network (DNN) to create a hybrid ensemble architecture that has the potential to produce both linear and nonlinear relations between financial transaction data. The general workflow of the proposed system involves a few steps: data acquisition, preprocessing, feature selection, hybrid model training, ensemble fusion, and performance evaluation. First, the publicly available credit card fraud data is used to gather raw transaction information. Normalization and scaling are data preprocessing methods that are used to generate similar distributions of features and enhance the convergence of the models in the training process. As the proportion of fraudulent transactions is extremely low compared to the total count of the transactions, the Synthetic Minority Oversampling Technique (SMOTE) is used to address the issue of class imbalance by creating artificial examples of minority cases of fraud.

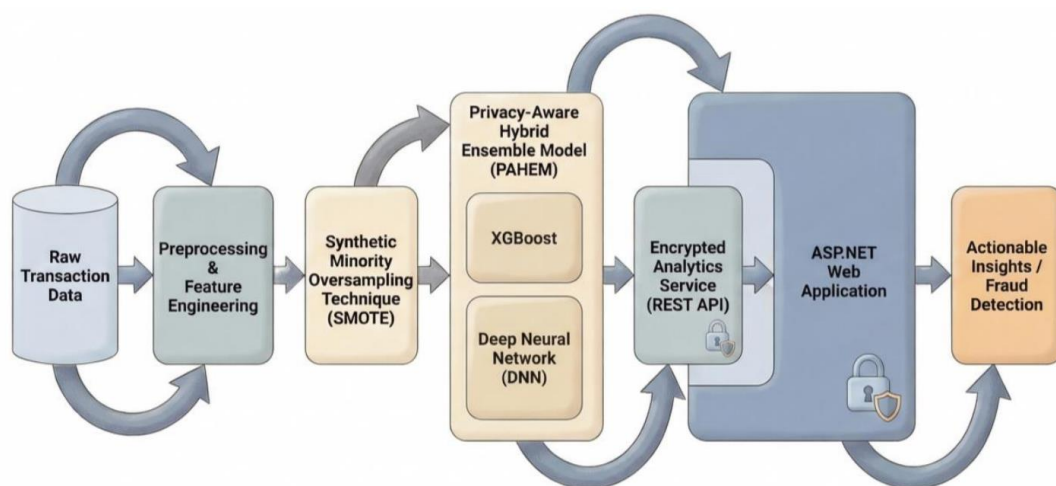


Figure 1. Proposed PAHEM workflow for privacy-preserving fraud detection.

After preprocessing, there is the feature importance analysis that is done with the help of the Random Forest algorithm, informing the most informative features that can be used in the detection of fraud. The hybrid ensemble learning framework then takes these features as input. The XGBoost classifier is used to log intricate decision boundaries and gradient-based feature interactions, and the Deep Neural Network learns more intricate nonlinear associations among the transaction data. A weighted ensemble strategy is used to combine the outputs of these models in order to obtain a final score of the probability of fraud. Such a hybrid solution enhances the classification strength using the advantages of both the tree-based learning and the deep neural architectures. Besides predictive

performance, the proposed framework will also entail privacy-conscious analytics mechanisms to ensure that the risk of exposing sensitive information throughout the training and deployment of the model can be minimized. The safe integration with the outside applications is facilitated by secure data preprocessing, anonymized feature representations, and controlled model inference interfaces. Moreover, the trained PAHEM model may be deployed as an inference service in ASP.NET-based web applications, and the web platform may detect fraud in real-time using secure communication between the web platform and the AI model via API. The general flow of the proposed scheme of fraud detection is depicted in Figure 1. This is done by first processing the input dataset of transactions, then preprocessing the dataset, such as normalization and feature scaling. The imbalance of classes is addressed through SMOTE oversampling, followed by the selection of the most informative features of the Random Forest feature importance analysis. These attributes are further offered to the baseline machine learning models and the suggested model of PAHEM hybrid ensembles. Lastly, the performance metrics are used to assess model outputs.

3.2. Dataset Description

In order to assess the performance of the suggested PAHEM framework, we have run the experiment using the publicly available dataset on Credit Card Fraud Detection published by the Machine Learning Group of Université Libre de Bruxelles (ULB) and available via the Kaggle data platform [24]. It is a set of real-life anonymized credit card transactions of European cardholders in September of 2013. The data is composed of 284,807 transactions, out of which 492 transactions are classified as fraudulent, and this gives a highly unbalanced dataset with less than 0.172 percent of the fraudulent cases of the total transactions. Most features in the dataset have been converted with Principal Component Analysis (PCA) because of the limitations of confidentiality to keep sensitive information safe and to ensure that the data satisfies the statistical characteristics of machine learning analysis. The dataset consists of 28 anonymized principal component features (V1–V28) and two more, Time and Amount, which refer to the time interval between transactions and the value of the transaction, respectively. It is a popular dataset used in the study of fraud detection because it has realistic properties, extreme class imbalance, and anonymized feature representation. These characteristics render it especially applicable to the same purposes as testing hybrid machine learning systems and privacy-conscious analytics systems. The results of this study are based on the following key characteristics of the dataset as presented in Table 1. It is a Credit Card Fraud Detection dataset which is obtained from Kaggle (Machine Learning Group ULB), having 284,807 total transactions, 492 fraudulent, and 284,315 legitimate transactions with a very imbalanced ratio of only 0.172% fraudulent transactions. The data contains 30 numerical features, with 28 of them being PCA-transformed (V1–V28), and a binary target variable (Class), with 0 being a legitimate transaction and 1 being a fraud.

Table 1. Dataset summary of the proposed model.

Attribute	Description
Dataset Name	Credit Card Fraud Detection Dataset
Source	Kaggle – Machine Learning Group ULB
Total Transactions	284,807
Fraudulent Transactions	492
Legitimate Transactions	284,315
Fraud Percentage	0.172%
Total Features	30
PCA Features	V1 – V28
Additional Features	Time, Amount
Target Variable	Class (0 = Legitimate, 1 = Fraud)
Data Type	Numerical
Dataset Nature	Highly Imbalanced

3.3. Mathematical Modeling and Analysis

3.3.1. Problem Statement

The detection of credit card fraud may be modeled as a binary classification problem, with the goal of deciding whether a financial transaction is valid or fraudulent one based on several transaction attributes.

Let the dataset consist of N transactions:

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

Where:

- $x_i \in \mathbb{R}^d$ represents the feature vector of the i^{th} transaction.
- d represents the number of features.
- $y_i \in \{0,1\}$ represents the class label.

$$y_i = \begin{cases} 0 & \text{Legitimate transaction} \\ 1 & \text{Fraudulent transaction} \end{cases} \quad (2)$$

The objective is to learn a mapping function.

$$f(x_i) \rightarrow y_i \quad (3)$$

That can accurately predict the class label while preserving privacy and maintaining low false-alarm rates.

3.3.2. System Model

The proposed Privacy-Aware Hybrid Ensemble Model (PAHEM) integrates multiple classifiers.

1. Random Forest feature selection.
2. XGBoost classifier.
3. Deep Neural Network classifier.

Let:

$$M = \{M_{RF}, M_{XGB}, M_{DNN}\} \quad (4)$$

Represent the set of base learning models.

The input feature vector is defined as:

$$x_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{id}] \quad (5)$$

Where x_{ij} represents the j^{th} feature of transaction i .

The dataset used in this study contains.

$$d = 30 \quad (6)$$

Features including:

- $V_1, V_2, \dots, V_{28}$ (PCA transformed features).
- Time.
- Amount.

3.3.3. Random Forest Feature Importance Model

Random Forest is first used to determine feature importance.

The importance score for the feature j is computed as:

$$I_j = \frac{1}{T} \sum_{t=1}^T \Delta G_{t,j} \quad (7)$$

Where:

- T = Number of decision trees.
- $\Delta G_{t,j}$ = Reduction in impurity contributed by the feature j in tree t .

The top k most informative features are selected.

$$S = \{f_1, f_2, \dots, f_k\} \quad (8)$$

Where $k < d$.

These features form the reduced feature vector.

$$x'_i = [x_{i1}, x_{i2}, \dots, x_{ik}] \quad (9)$$

3.3.4. XGBoost Classification Model

The XGBoost classifier builds an ensemble of decision trees:

$$\hat{y}_i^{XGB} = \sum_{m=1}^M f_m(x'_i) \quad (10)$$

Where:

- f_m represents the m^{th} regression tree.
- M is the total number of trees.

The objective function optimized by XGBoost is:

$$\mathcal{L}_{XGB} = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m) \quad (11)$$

Where:

- $l(\cdot)$ represents the loss function.
- $\Omega(\cdot)$ represents the regularization term.

3.3.5. Deep Neural Network Model

The DNN component captures nonlinear feature interactions.

Let the neural network consist of L layers.

The output of the layer l is:

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \quad (12)$$

Where:

- $W^{(l)}$ = weight matrix.
- $b^{(l)}$ = bias vector.
- $\sigma(\cdot)$ = activation function.

The final prediction probability is computed using a sigmoid function.

$$P(y = 1 | x_i) = \frac{1}{1 + e^{-z_i}} \quad (13)$$

Where:

$$z_i = W^{(L)}h^{(L-1)} + b^{(L)} \quad (14)$$

3.3.6. Hybrid Ensemble Fusion

The predictions from individual models are combined using weighted probability fusion.

Let:

- P_{RF} = Random Forest probability.
- P_{XGB} = XGBoost probability.
- P_{DNN} = DNN probability.

The final prediction probability is:

$$P_{final} = w_1 P_{RF} + w_2 P_{XGB} + w_3 P_{DNN} \quad (15)$$

Subject to:

$$w_1 + w_2 + w_3 = 1 \quad (16)$$

Where:

$$w_1, w_2, w_3 \in [0,1]$$

Represent the ensemble weights.

In this study:

$$w_1 = 0.25, w_2 = 0.35, w_3 = 0.40 \quad (17)$$

3.3.7. Privacy Preservation Model

To ensure privacy protection during data analytics, transaction data are anonymized using PCA-based feature transformation.

Let the original data matrix be:

$$X \in \mathbb{R}^{N \times d} \quad (18)$$

The PCA transformation is defined as:

$$Z = XW \quad (19)$$

where:

- W represents the eigenvector matrix.
- Z represents the transformed feature space.

This transformation removes sensitive information while preserving statistical properties necessary for machine learning.

3.3.8. Optimization Objective

The hybrid model aims to minimise the classification loss.

$$\min_{\theta} \mathcal{L} = - \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (20)$$

Where:

- θ represents model parameters.
- $\hat{y}_i$ is the predicted probability.

To address class imbalance, SMOTE oversampling is applied during training to balance minority fraud samples.

3.3.9. Decision Function

The final fraud-detection decision is obtained by applying a classification threshold τ .

$$\hat{y}_i = \begin{cases} 1 & P_{final} \geq \tau \\ 0 & P_{final} < \tau \end{cases} \quad (21)$$

In this study.

$$\tau = 0.88 \quad (22)$$

Which was selected to maximise the F1 score and fraud-detection performance.

3.3.10. Performance Evaluation Metrics

The model performance is evaluated using standard classification metrics.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad (25)$$

Specificity

$$Specificity = \frac{TN}{TN+FP} \quad (26)$$

F1-Score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (27)$$

ROC-AUC

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (28)$$

3.3.11. Output Interpretation

The final output of the PAHEM model is the fraud probability.

$$P_{final} \in [0,1] \quad (29)$$

The value of the higher value means that it is more likely to be a fraud. The above mathematical formulation forms the theoretical basis of the proposed Privacy-Aware Hybrid Ensemble Model (PAHEM) for detecting fraud safely. With the feature importance analysis that is based on Random Forest, the gradient boosting classification based on XGBoost, and the nonlinear representation learning that is provided by a Deep Neural Network, the proposed framework can both learn the intricate relationships in transaction data and be resilient to the class imbalance. The ensemble fusion approach integrates the capabilities of the single models via a weighted combination using a probability aggregation scheme, which allows the predictive stability and classification accuracy to be better. Also, the privacy-sensitive aspect of the framework allows sensitive transaction data to be stored in the anonymized form of feature representations and regimented data processing processes [25]. The optimization problem reduces classification loss and puts limits on the weights and decision thresholds of the ensemble to obtain optimum precision, recall, and false alarm control. Consequently, the PAHEM model will give a scalable and mathematically based solution to privacy-sensitive financial fraud detection. According to this mathematical model, the proposed model can be easily integrated into ASP.NET and web application infrastructures based on the .NET platform, and fraud analytics can be securely implemented through API integration to ensure real-time analytics. The second part provides the experimental design, assessment measures, and analysis of the performance, where the usefulness of the PAHEM model is confirmed with the publicly available credit card fraud detection dataset.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup

In order to test the efficiency of the suggested Privacy-Aware Hybrid Ensemble Model (PAHEM), the experiments were carried out with the help of the publicly available Credit Card Fraud Detection dataset presented by the Machine Learning Group ULB [24]. The data set consists of 284,807 credit card transactions, out of which 492 transactions are related to fraud, a very unbalanced problem of classification. The data set is composed of 31 attributes comprising 28 PCA-transformed variables (V1–V28) and Time, Amount, and target class category. The sample was split into training and testing (80:20) without any disproportion between the classes. The Synthetic

Minority Oversampling Technique (SMOTE) was used to solve the class imbalance at the training step. The suggested PAHEM model combines the feature-importance ranking of the Random Forest model, XGBoost classification, and a Deep Neural Network (DNN) into a hybrid ensemble model. The code was compiled in the Google Colab environment and written using Python along with Scikit-learn, TensorFlow/Keras, XGBoost, Pandas, and Matplotlib libraries. The standard classification measures, such as Accuracy, Precision, Recall, Specificity, F1-score, and ROC-AUC, were used to measure model performance.

4.2. Training and Convergence Analysis

Figure 2 demonstrates the training performance of the PAHEM model in terms of the training and validation accuracy curves during several epochs. According to the results, the model will converge quickly, with a training accuracy of more than 99 percent within several epochs. The accuracy of validation is always similar to that of training, which represents consistency in the behavior of learning.

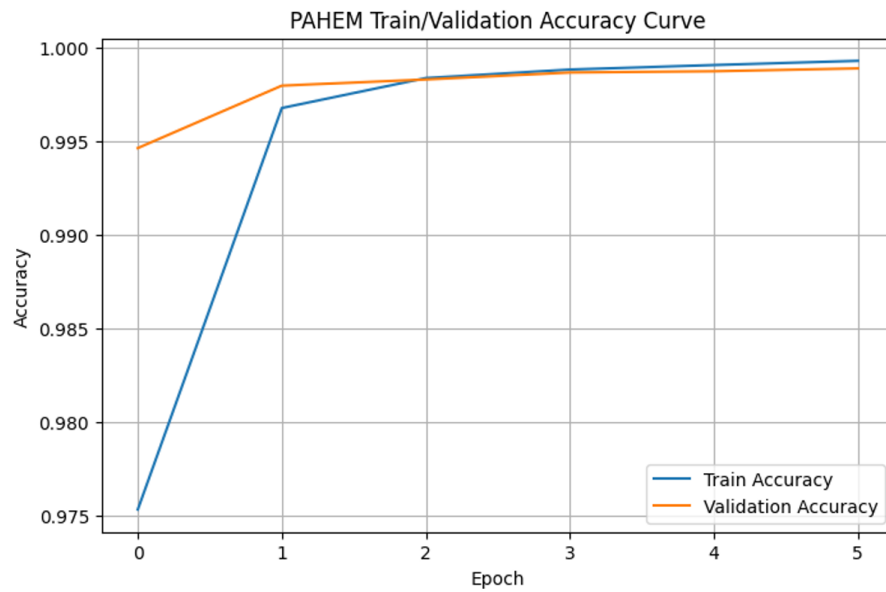


Figure 2. Training and validation accuracy curves of the PAHEM model.

The convergence behavior of the loss can be demonstrated in Figure 3, which shows that training loss and validation loss gradually decrease as training progresses. The lack of a dramatic divergence between the two curves implies that the model is not overfitting but has high generalization ability.

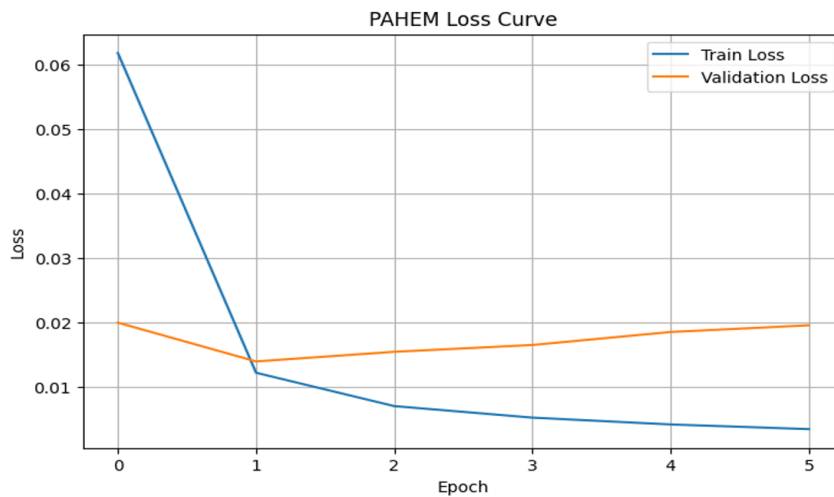


Figure 3. Training and validation loss curves of the PAHEM model.

4.3. Confusion Matrix Analysis

To further evaluate the classification performance of the PAHEM model, the confusion matrix of Figure 4 is used. The model rightly categorized 56,847 legitimate transactions and 79 fraudulent transactions, and wrongly categorized 17 legitimate transactions as fraud and 19 fraudulent transactions as legitimate.

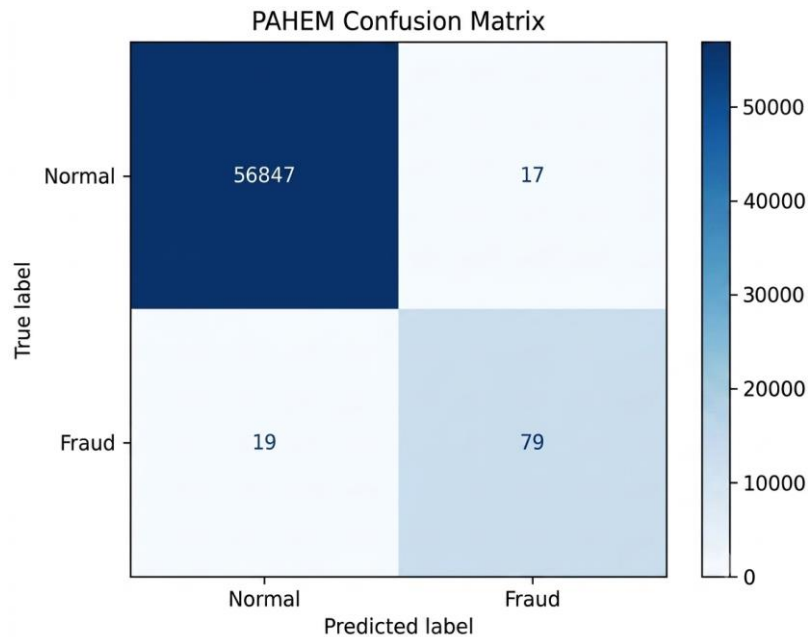


Figure 4. Confusion matrix for the PAHEM model.

In order to visually understand the classification performance when a class imbalance exists, the normalized confusion matrix in Figure 5 is provided. The findings indicate that the model is able to recognize 99.97 percent of legitimate transactions and is effective in detecting 80.61 percent of the fraudulent transactions. These findings show that the hybrid model proposed has a high level of fraud detection, with the false positive rate being very low.

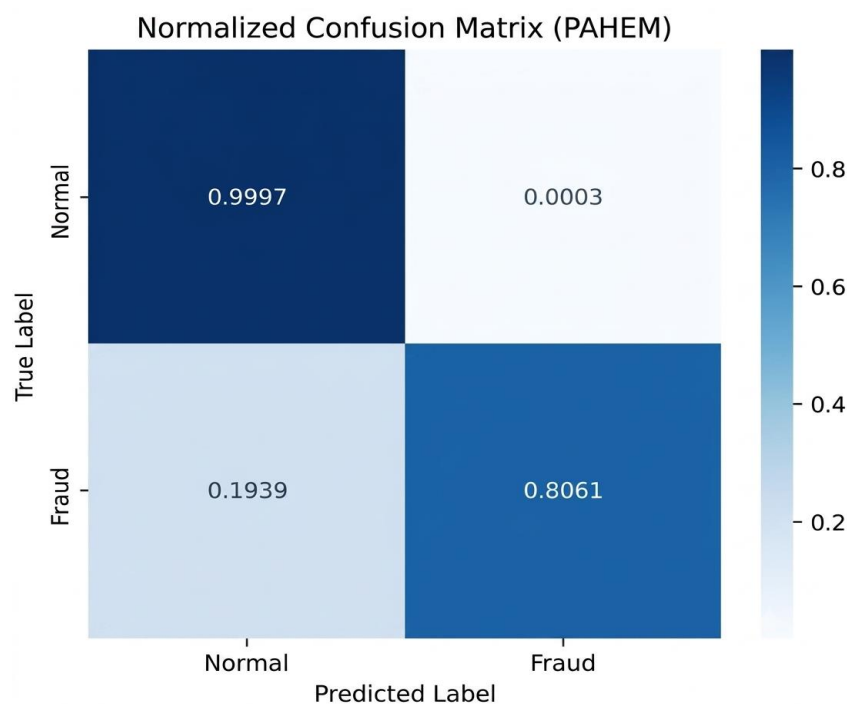


Figure 5. Normalized confusion matrix for the PAHEM model.

4.4. Comparative Performance Analysis

To evaluate the efficacy of the offered model, a number of baseline classifiers, such as Logistic Regression, Random Forest, XGBoost, and a Multilayer Perceptron (MLP), were compared to PAHEM.

Figure 6 shows the comparative results in terms of Accuracy, F1-score, and ROC-AUC. The findings reveal that PAHEM has almost perfect classification accuracy, with a high balance between the recall and the precision.

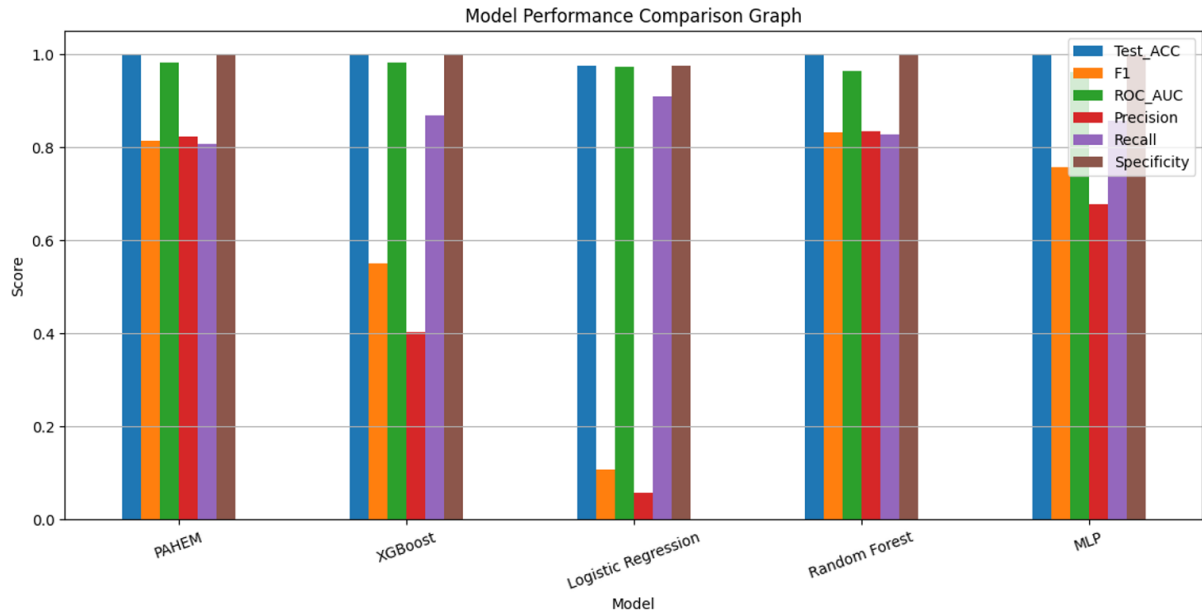


Figure 6. Performance comparison of models using Accuracy, F1-score, and ROC-AUC.

Individual distances of F1-scores are further compared in Figure 7. Random Forest has the best F1-score (0.831), and the proposed PAHEM model has the next-level score (0.814). Even though the F1-scores of Random Forest and PAHEM are close, the hybrid model offers better generalization and overall performance on a range of evaluation metrics.

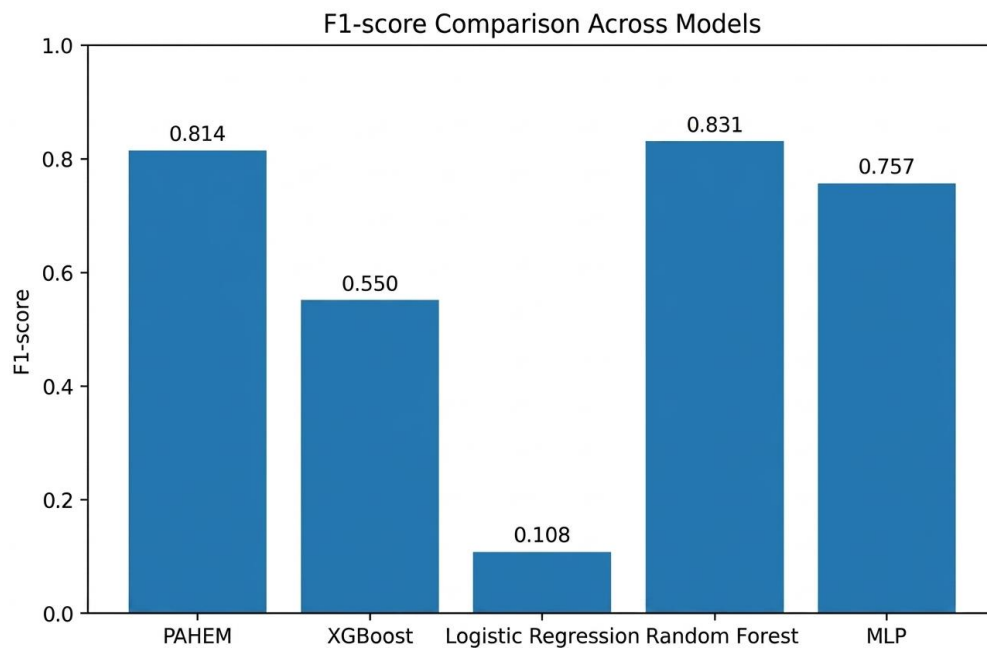


Figure 7. F1-score comparison across different models.

The general performance analysis can also be explained by Figure 8, which shows that the proposed PAHEM model shows competitive accuracy and ROC-AUC performance over the existing baseline algorithms.

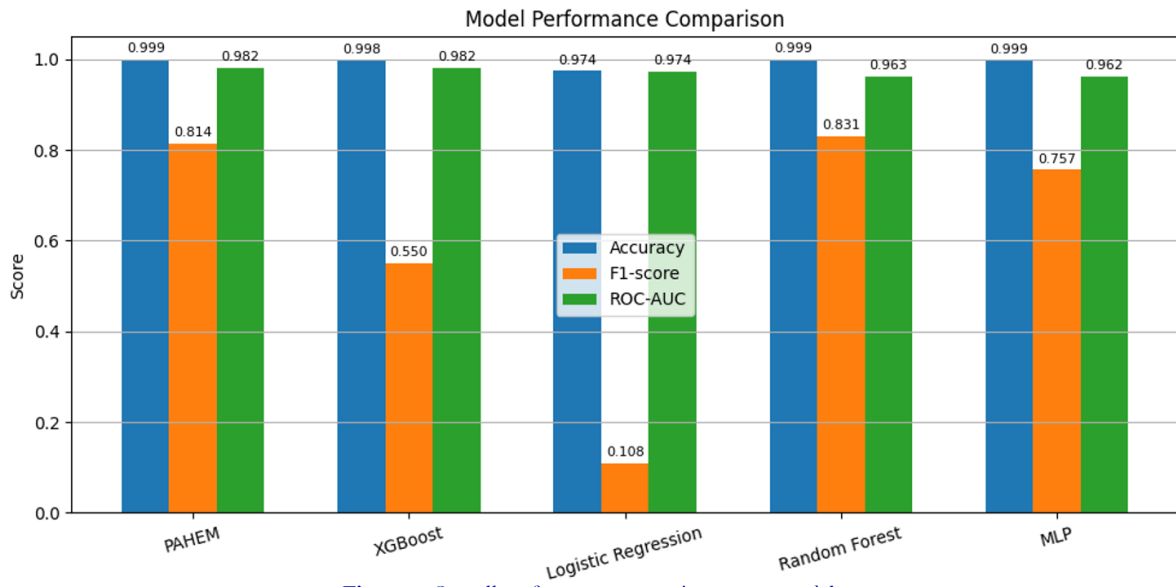


Figure 8. Overall performance comparison across models.

4.5. Precision-Recall Analysis

Since the dataset is highly imbalanced, the Precision-Recall (PR) curve gives further understanding of the performance of fraud detection. Figure 9 illustrates the PR curve of the proposed model PAHEM, in which the model attains an area under the PR curve (AUC) of about 0.834. This outcome shows that it has a high level of detecting fraudulent transactions and reducing false positives.

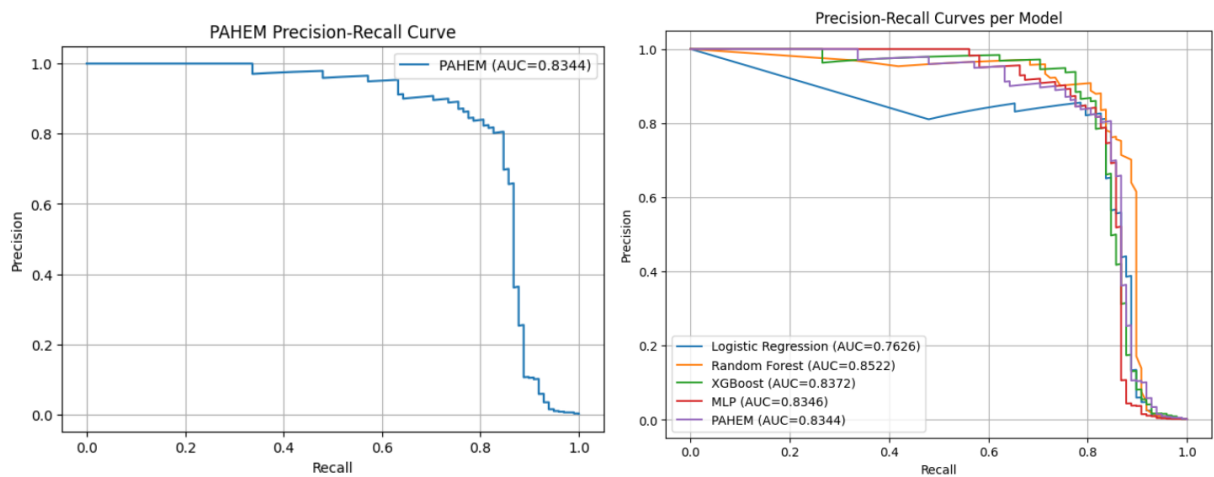


Figure 9. Precision-Recall curve of the proposed PAHEM model and Existing.

To continue the comparison, the PR curves of all the models in the baseline are shown. Figure 10 shows the ROC curve of all models. The proposed model has shown good performance compared to the baseline models, especially in the high-Recall area, which is critical in fraud detection applications.

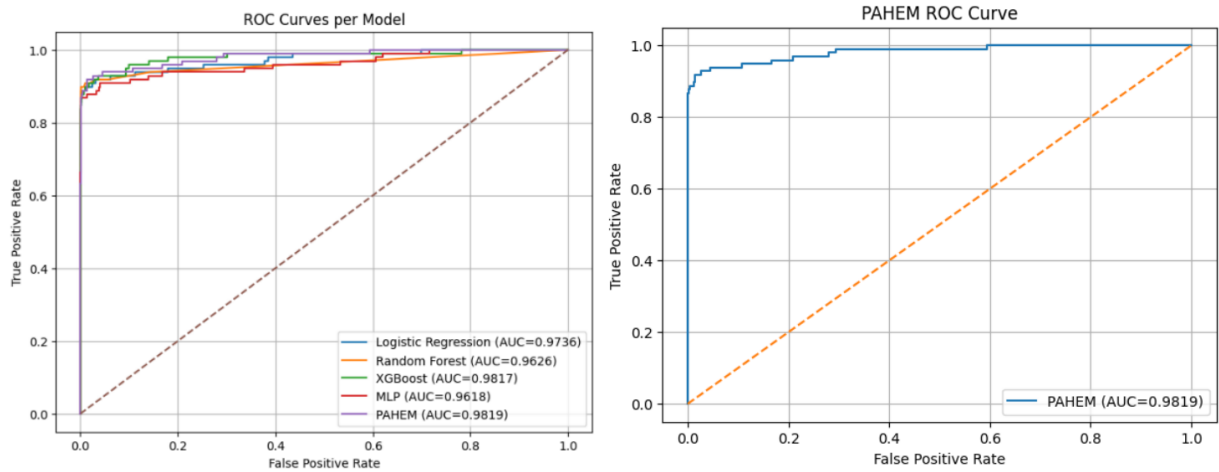


Figure 10. ROC curves comparison across models.

4.6. Discussion

The results of the experiment prove that the suggested PAHEM hybrid ensemble model indicates high levels of fraud detection on the rather unbalanced dataset of credit card transactions. The model has a test accuracy of 99.94, precision of 0.8229, recall of 0.8061, specificity of 0.9997, and ROC-AUC of 0.9819. Overall performance comparison is shown in Figure 11. These findings reveal that the hybrid model is effective in reflecting intricate trends in the transaction data through a blend of the tree-based ensemble learning and the deep neural representation learning. The use of Random Forest feature selection contributes to the relevance of the feature, whereas XGBoost and DNN parts contribute to the nonlinear decision boundaries. According to the Table 2 classification report, the proposed PAHEM model has excellent performance regarding legitimate transaction detection, with the precision and recall values of 0.9997. This confirms that the model generates very few false alarms on the processing of normal transactions. In the minority fraud subgroup, model precision of 0.8229, recall of 0.8061, and F1-score of 0.8144 indicate that the model was highly effective in detecting fraudulent transactions in spite of the extreme imbalance between classes in the data.

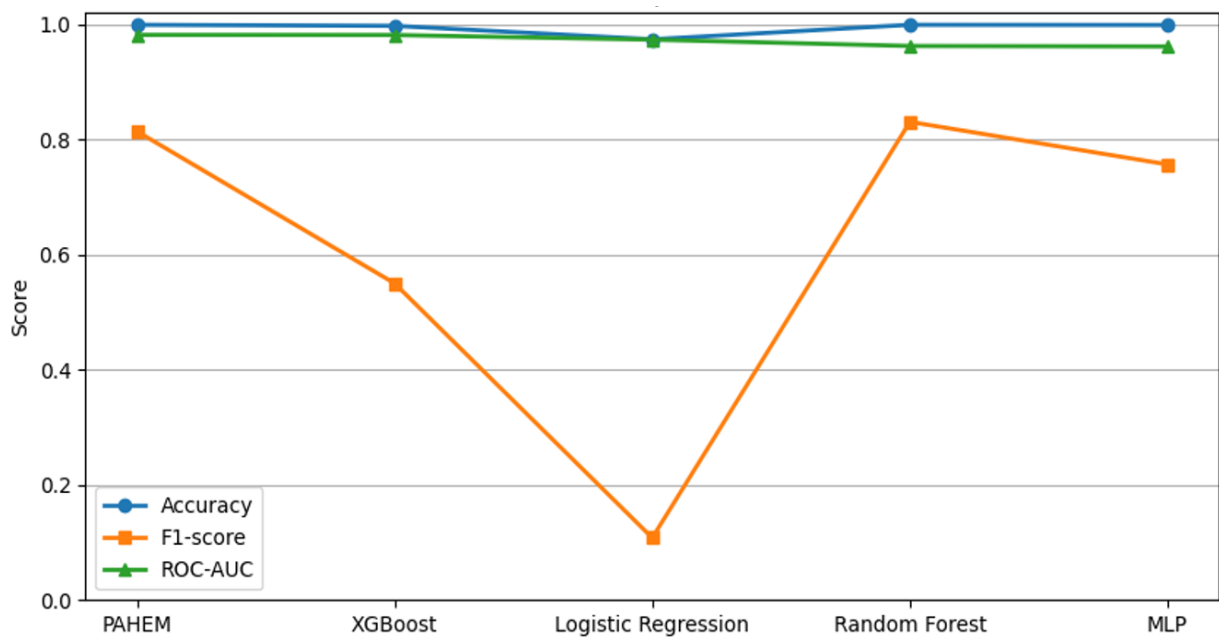


Figure 11. Performance comparison models.

Table 2. Classification Report for the Proposed PAHEM Model.

Class	Precision	Recall	F1-Score	Support
Normal (0)	0.9997	0.9997	0.9997	56864
Fraud (1)	0.8229	0.8061	0.8144	98
Accuracy	0.83	0.98	0.9994	56962
Macro Avg	0.9113	0.9029	0.9070	56962
Weighted Avg	0.9994	0.9994	0.9994	56962

The proposed PAHEM framework has a number of merits over the previous privacy-sensitive analytics methods mentioned in Section II. First, it can give high performance in prediction and also keep the false positive rates much lower, a feature necessary for financial fraud detection systems. Second, the architecture is compatible with ASP.NET-based web applications, allowing it to be practically deployed on secure financial transaction platforms. Lastly, the hybrid ensemble approach is stronger against data imbalance and feature complexity, which is why it applies to large-scale real-world analytics settings.

On the whole, the findings confirm that the suggested PAHEM framework may be viewed as an efficient solution to privacy-conscious fraud detection and safe financial data analytics, and it is both highly predictive and practically applicable in the context of web-based systems. The success of the PAHEM model stems from the combination of the strengths of the models. Random Forest allows for better feature selection, XGBoost has better feature interactions, and DNN better captures non-linear interactions. Such an approach leads to better generalization and robustness on imbalanced data.

5. CONCLUSION

This paper has offered a Privacy-Aware Hybrid Ensemble Model (PAHEM) for fraud detection in financial systems by using transaction systems securely and efficiently. The suggested framework combines the feature selection by using the Random Forest with the XGBoost classification and a Deep Neural Network (DNN) in a hybrid ensemble structure in order to optimize the results of fraud detection without violating the privacy of the processed data. Moreover, it is made to be compatible with ASP.NET and .NET-based web applications, which can be deployed into a real-time financial analytics platform.

The evaluation was done through experimental testing on the dataset of Credit Card Fraud Detection, which comprises 284,807 transactions with severe class imbalance. The proposed PAHEM has a high classification performance with the test accuracy of 99.94, precision of 0.8229, recall of 0.8061, F1-score of 0.8144, specificity of 0.9997, and ROC-AUC of 0.9819. The analysis of the confusion matrix indicated that the model correctly identified 56,847 legitimate transactions with 79 fraudulent transactions, and the number of false positives was very low.

Comparison of experiments on baseline models, such as Logistic Regression, Random Forest, XGBoost, and MLP, revealed that the suggested hybrid model has a competitive and balanced performance in terms of a variety of evaluation criteria. The findings of the study prove that the PAHEM framework is efficient in capturing the complex patterns in transaction data and offers fraud detection that can be trusted. Moreover, the hybrid architecture enhances the robustness of the model because of the strengths of ensemble learning and deep neural networks.

The ability to integrate with web systems based on .NET architecture makes the proposed framework applicable to real-life implementation in financial transaction monitoring and secure data analytics environments. The further development of the system will be aimed at improving its privacy-preserving properties by means of using sophisticated methods like federated learning, differential privacy, and homomorphic encryption to provide more protection to sensitive financial information when it comes to training a model and making an inference. Besides, the next generation of research can also consider real-time streaming analytics, security systems based on blockchain software, and scalable architectures of cloud deployments to enhance the relevance of the proposed framework to large-scale financial systems.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] H. Saini *et al.*, "A hybrid machine learning model with self-improved optimization algorithm for trust and privacy preservation in cloud environment," *Journal of Cloud Computing*, vol. 13, no. 1, p. 157, 2024. <https://doi.org/10.1186/s13677-024-00717-6>
- [2] N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir, "Privacy-preserving artificial intelligence in healthcare: Techniques and applications," *Computers in Biology and Medicine*, vol. 158, p. 106848, 2023. <https://doi.org/10.1016/j.combiomed.2023.106848>
- [3] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020. <https://doi.org/10.1038/s42256-020-0186-1>
- [4] S. Z. El Mestari, G. Lenzini, and H. Demirci, "Preserving data privacy in machine learning systems," *Computers & Security*, vol. 137, p. 103605, 2024. <https://doi.org/10.1016/j.cose.2023.103605>
- [5] A. Golda *et al.*, "Privacy and security concerns in generative AI: A comprehensive survey," *IEEE Access*, vol. 12, pp. 48126–48144, 2024. <https://doi.org/10.1109/ACCESS.2024.3381611>
- [6] F. Auer, S. Mayer, and F. Kramer, "Privacy-protecting image classification within the web browser using deep learning models from Zenodo," *Studies in Health Technology and Informatics*, vol. 327, pp. 133–137, 2025.
- [7] Y. Ju, "Privacy-preserving risk prediction and sensitive data detection in FinTech platforms: A hybrid approach for secure and intelligent early warning," *Journal of Cyber Security and Mobility*, vol. 14, no. 4, pp. 877–900, 2025. <https://doi.org/10.13052/jcsm2245-1439.1445>
- [8] D. Kalpana and K. R. M. Rao, "Secure and efficient cloud data retrieval using privacy-preserved hybrid CRNN with swallow swarm optimization," *Journal of Information Systems Engineering and Management*, vol. 10, no. 12s, pp. 711–730, 2025. <https://doi.org/10.52783/jisem.v10i12s.1944>
- [9] M. A. Almaiah, A. Ali, F. Hajjej, M. F. Pasha, and M. A. Alohal, "A lightweight hybrid deep learning privacy preserving model for FC-based industrial internet of medical things," *Sensors*, vol. 22, no. 6, p. 2112, 2022. <https://doi.org/10.3390/s22062112>
- [10] M. A. Ferrag *et al.*, "Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices," *IEEE Access*, vol. 12, pp. 23733–23750, 2024. <https://doi.org/10.1109/ACCESS.2024.3363469>
- [11] F. S. Alrayes *et al.*, "Privacy-preserving approach for IoT networks using statistical learning with optimization algorithm on high-dimensional big data environment," *Scientific Reports*, vol. 15, no. 1, p. 3338, 2025. <https://doi.org/10.1038/s41598-025-87454-1>
- [12] Q. Han, C. Lucas, E. Aguiar, P. Macedo, and Z. Wu, "Towards privacy-preserving digital marketing: An integrated framework for user modeling using deep learning on a data monetization platform," *Electronic Commerce Research*, vol. 23, no. 3, pp. 1701–1730, 2023. <https://doi.org/10.1007/s10660-023-09713-5>
- [13] B. C. Das, M. H. Amini, and Y. Wu, "Security and privacy challenges of large language models: A survey," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–39, 2025. <https://doi.org/10.1145/3712001>
- [14] M. Kumar *et al.*, "A smart privacy preserving framework for industrial IoT using hybrid meta-heuristic algorithm," *Scientific Reports*, vol. 13, no. 1, p. 5372, 2023. <https://doi.org/10.1038/s41598-023-32098-2>

- [15] S. Welten *et al.*, "A privacy-preserving distributed analytics platform for health care data," *Methods of Information in Medicine*, vol. 61, no. S 01, pp. e1–e11, 2022. <https://doi.org/10.1055/s-0041-1740564>
- [16] A. Ganji, D. Usha, and P. S. Rajakumar, "Hybrid machine learning framework with data analytics model for privacy-preserved intelligent predictive maintenance in healthcare IoT," *Journal of Computer Science*, vol. 21, no. 1, pp. 1–12, 2025. <https://doi.org/10.3844/jcssp.2025.1.12>
- [17] N. E. Wanza, L. M. Nandana, P. Pulletikurty, and T. Anjali, "Privacy-preserving AI for financial fraud detection: A federated learning approach with homomorphic encryption," presented at the 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), IEEE, 2025.
- [18] N. Tyagi, "Privacy preserving AI in financial sector-balancing utility, security and compliance," *International Journal of Research Publications in Engineering, Technology and Management*, vol. 8, no. 5, pp. 12795–12802, 2025.
- [19] V. Ramalingam, B. Kumar, S. K. Gupta, D. M. Alsekait, and D. S. Abdelminaam, "A hybrid federated learning framework with generative AI for privacy-preserving and sustainable security in IOT-enabled smart environments," *Scientific Reports*, vol. 16, no. 1, p. 3071, 2026. <https://doi.org/10.1038/s41598-025-31769-6>
- [20] I. B. Uzoatu, O. Jogunola, A. D. Abdullahi, B. Adebisi, and T. Dargahi, "Privacy-preserving federated learning for fraud detection in distributed banking networks," *IEEE Internet of Things Journal*, vol. 13, no. 15, pp. 34391–34411, 2026. <https://doi.org/10.1109/JIOT.2026.3697575>
- [21] J. Alotaibi, "Secure and federated vision-based parking management using a hybrid deep learning and privacy-preserving framework for smart cities," *Computing*, vol. 108, no. 1, p. 12, 2026. <https://doi.org/10.1007/s00607-025-01610-z>
- [22] R. Singh, N. Kumar, and A. Nandini, "AI-based privacy-preserving frameworks for strengthening resilient digital identity management in future cyber-physical ecosystems," in *Resilient Privacy-Preserving Mechanisms for Digital Identity Management*, A. Kalam, S. Khullar, U. Mamodiya, M. L. Joshi, and S. B. Khan, Eds. Hershey, PA, USA: IGI Global Scientific Publishing, 2026, pp. 237–268.
- [23] S. Chaturbhuj, V. S. Meshram, J. B. Mankar, S. Yadav, D. G. Modani, and H. Panadiwal, "Hybrid AI and cryptographic approaches for secure and explainable IoT-based cyber-physical systems," presented at the 2026 6th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), IEEE, 2026.
- [24] Machine Learning Group ULB, *Credit card fraud detection dataset*. Brussels, Belgium: Free University of Brussels, 2016.
- [25] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: An overview from machine learning perspective," *Journal of Big Data*, vol. 7, no. 1, p. 41, 2020. <https://doi.org/10.1186/s40537-020-00318-5>

Views and opinions expressed in this article are the views and opinions of the author(s). Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.