Acta Informatica Pragensia
2027, Volume 16, Issue 1, Forthcoming article
https://doi.org/10.18267/j.aip.327

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Article Open Access

Enhancing Fraud Detection Performance on Imbalanced Datasets: A ResNet-Based Approach Integrating Multi-Attention Techniques and SMOTE-ENN

Mohammed Khaldoon Altalib [1], **Karam Abdullah** [1], **Naomie Salim** [2]

[1] Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Nineveh, Iraq
[2] UTM Big Data Center, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Corresponding author: Mohammed Khaldoon Altalib (mohammadaltalib79@uomosul.edu.iq)

Editorial Record

First submission received:
January 4, 2026

Revisions received:
April 2, 2026
June 2, 2026
July 21, 2026

Accepted for publication:
August 8, 2026

Academic Editor:
Adam Borovicka
Prague University of Economics and Business, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Altalib, M. K., Abdullah, K., & Salim, N. (2027). Enhancing Fraud Detection Performance on Imbalanced Datasets: A ResNet-Based Approach Integrating Multi-Attention Techniques and SMOTE-ENN. *Acta Informatica Pragensia*, 16(1), Forthcoming article. https://doi.org/10.18267/j.aip.327

Copyright:
© 2027 by the author(s). Licensee Prague University of Economics and Business, Czech Republic. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

Abstract

Background: The detection of financial fraud has become increasingly significant because electronic transactions and advanced frauds have been growing rapidly, leading to huge losses for the economy. Yet, significant class imbalance, data heterogeneity, evolving fraud trends, and a small number of labelled samples remain as challenges for building effective fraud detection systems.

Objective: This article assesses an integrated approach to a fraud-detection framework based on residual networks (ResNet), different attention mechanisms, and class-imbalance handling techniques across heterogeneous financial datasets.

Methods: The proposed experimental framework combined ResNet with self-attention, channel attention, gated attention and multihead attention. Class imbalance was addressed using the synthetic minority oversampling technique (SMOTE) and edited nearest neighbours (ENN). The model was evaluated on the European Credit Card Fraud dataset, the UCI Taiwan dataset, and the combined dataset to assess its performance across both transactional and demographic feature spaces.

Results: On the European Credit Card Fraud dataset, the ResNet multihead attention model outperformed all other models, while on the UCI Taiwan dataset, the ResNet gated-attention model had the best F1 performance. The channel-attention model performed the best in terms of accuracy and AUC for the merged data. AUC was not always the best choice for the various experimental setups, but the most consistent improvements were in F1 score and ranking stability.

Conclusion: The integrated ResNet-attention framework achieved competitive performance on fraud detection in imbalanced and heterogeneous financial datasets, particularly in terms of F1 score. The results show that the utility of single attention mechanisms varies across datasets and does not indicate that any single architecture is superior. The framework should be tested on other real-world financial datasets, and further optimisation should be explored to enhance its generalisability and robustness.

Index Terms

Deep learning; ResNet; Attention mechanisms; Class imbalance; Synthetic minority oversampling technique; Edited nearest neighbours; Data merging.

1 INTRODUCTION

The Federal Trade Commission, according to its recent findings, indicates that due to exponential growth in electronic transactions, there has been a significant growth in losses due to fraud. Fraudulent activity in 2024 alone resulted in consumers losing more than 12.5 billion USD, which, supported by the increased number of fraud cases, highlights the growing financial costs associated with cyber fraud throughout the economy.

These illicit transactions not only result in huge financial losses, but also affect consumer confidence and economic shocks. The tendency of individuals to engage in legitimate business is minimized when individuals think that online platforms are not safe, thus the market forces in general are not trusted. The adequate fraud detection systems can serve two purposes: they can not only assist in minimizing the losses but also contribute to the growth of confidence in online business. These systems will mitigate risk exposures, build consumer confidence and achieve stability of the electronic financial infrastructures overall by correctly detecting and intercepting unauthorised transactions. However, intrinsic difficulties, namely, the inherent imbalance in the data, addressing the continuously changing nature of the fraud trends and the lack of labelled data, severely complicate the development of strong and reliable detection models (Ali et al., 2022; Lindemulder, 2024).

Deep learning has become an effective approach for fraud detection because it can identify complex patterns in large-scale financial data as it can spot complex patterns in large datasets. In various models out there, the residual network (ResNet) stands out. It uses skip connections to deal with the challenges of the deep network, which helps find important features without losing performance. This is important in detecting fraud, where small irregularities may give suspicious indications. Learning models can be promoted by adding attention mechanisms. These allow the models to focus on the most important parts of the data, making them easier to understand and making prediction more accurate. In detecting fraud, where transactions are complex, techniques such as self-attention, channel attention and multihead attention help the model distinguish between fake and real transactions, which matters the most (Cao et al., 2021; Huang et al., 2025; Mienye & Jere, 2024; Saeed et al., 2024; Younis et al., 2025).

Handling unbalanced data is an important issue in detecting fraud, as fraud cases usually represent a very small fraction of transactions, leading to biased predictions in favour of the majority classes. Synthetic minority over-sampling technique (SMOTE) and edited nearest neighbours (ENN) techniques that are used to address unbalanced data. SMOTE produces new synthetic samples by interpolating between existing minority class samples, while ENN removes noise and irrelevant samples from the majority class. This dual approach increases the strength of the model, ensures improvement in generalisation and reduces false positive rates in classification functions. This improves classification performance on unbalanced data (Azhar et al., 2022; Bonde & Bichanga, 2025; Bounab et al., 2024).

Taking into consideration how models are trained in the context of fraud detection, merging datasets from different sources covers almost all behavioural patterns of transactions. Cross-dataset merging improves the breadth and depth of the training data, which supports models in learning different types of transactions as well as fraudulent activities. Moreover, systematic dataset merging allows a better assessment of how effective a certain model is and the overall adaptability of the system to different scenarios, which leads to improved performance in preventing fraudulent activities (Nguyen et al., 2023).

This study presents an integrated experimental framework that combines an established one-dimensional ResNet backbone with four attention mechanisms: self-attention, channel attention, gated attention units and multihead attention. SMOTE-ENN is incorporated into the training pipeline to address class imbalance. The study does not introduce a fundamentally new neural architecture; instead, it provides a controlled comparative evaluation in which the attention mechanism is varied while the main preprocessing, training and evaluation settings are maintained consistently. The framework is assessed on the European Credit Card dataset, the UCI Taiwan dataset and an exploratory mixed-dataset to investigate how the evaluated configurations behave under different feature distributions and class imbalance conditions.

The key contributions of the present research are as follows:

1. The study develops an integrated experimental framework that combines a ResNet backbone, multiple attention mechanisms and class imbalance handling techniques within a unified fraud detection pipeline.
2. The study provides a controlled comparative evaluation of multiple attention mechanisms within the same ResNet-based setting, enabling a clearer assessment of how different attention designs influence fraud detection performance under highly imbalanced conditions.

It evaluates the framework across the European Credit Card dataset, the UCI Taiwan dataset and an exploratory heterogeneous mixed setting to examine robustness under different financial data characteristics and class distributions.

1.1 Literature review

The recent years have witnessed an increase in the attention paid to financial fraud detection due to digital payments. Online banking, electronic financial services and payments have grown at a high rate. As fraud techniques are becoming more advanced, classical rule-based systems tend to be inadequate, because they find it difficult to discover intricate and dynamic patterns of fraudulent conduct. For this reason, machine learning and deep learning methods have been used in numerous recent studies to enhance better detection, lower false alarms and create stronger fraud detection systems. Overall, the past research in the field can be divided into three primary themes: (i) managing class imbalance, (ii) deep learning and hybrid architecture development, and (iii) model evaluation across varying kinds of financial data (Chen et al., 2025).

Fraud detection techniques face class imbalance issues across data classification problems. In financial datasets, there are far more legitimate transactions than fraudulent. Most classification models will learn from the data imbalance and learn almost exclusively from the majority class, failing to learn from the other class and thus not detecting fraudulent transactions. Various studies have offered solutions to issues surrounding class imbalance through the use of resampling techniques. In this regard, Esenogho et al. (2022) contributed to the class imbalance solution by building an ensemble of neural networks and applying SMOTE-ENN to increase the representation of the minority fraudulent samples. In a similar direction, Mienye and Sun (2023) presented a deep learning-based stacking ensemble model based on LSTM and GRU as base learners and MLP as a meta-learner. The authors used the SMOTE-ENN method to increase the representation of the minority classes. Zhao and Guan (2023) combined CTGAN-based oversampling with a temporal convolutional network and then further refined the training data with the help of the neighbourhood cleaning rule. Khalid et al. (2024) suggested a combination of KNN, SVM, bagging, random forest and boosting classifiers for their real-time detection to resolve the imbalanced data issue. The under-sampling and the SMOTE preprocessing techniques were applied and the European Credit Card dataset was trained on the model.

The second key line of the literature concentrates on deep learning and hybrid models of capturing the complex patterns of fraud. Subtle, nonlinear and multivariate fraudulent behaviour is the characteristic behaviour being investigated, and deep learning is especially appropriate in this context. Alarfaj et al. (2022) presented a comparative study involving both traditional machine learning (e.g., SVM, XGBoost, decision tree) and deep learning models, including CNN-based architectures. The paper sought to solve high false alarms and concept drift in fraud detection. An empirical analysis was made based on the European Credit Card dataset, which showed that CNN models with optimized hidden layers have a significant positive effect on performance. Jiang et al. (2023) proposed UAAD-FDNet, a framework of unsupervised anomaly detection, which involved the use of autoencoders, attention mechanisms and generative adversarial learning to identify fraud without relying on labelled data to a large extent. Almazroi and Ayub (2023) proposed a hybrid deep learning model to detect fraud as well and found it to be significantly better than traditional approaches. Similarly, Ileberi and Sun (2024) combined CNNs, LSTMs, transformers and XGBoost into a single framework and attempted to capture both spatial and sequential patterns in financial data related to credit. Combined, these papers indicate that deep learning models are increasingly useful in deriving meaningful representations of more complex financial data.

In this field of study, attention processes are of special significance. A model can use attention to give preference to portions of the input that need the most attention, and give less or no consideration to less relevant or noisy input. Whenever working on fraud detection, focus is especially crucial because in this field we might encounter weak and sparse or deep significant signals amidst normal transactions. Thus, models that are attention-based can provide better representation learning and better classification. As a result, a combination of deep learning models and attention-based systems is a highly beneficial research field to create a fraud detection system.

The other theme that is significant in the literature is application of various kinds of financial data. Although most studies consider transaction-level fraud detection, others look at customer-level credit behaviour, including default risk, repayment history and billing patterns. Examples include Emmanuel et al. (2024), where a credit risk prediction model was proposed as a stacked classifier and feature selection, and Bhandary and Ghosh (2025), who demonstrated that contemporary machine learning algorithms, and deep neural networks in particular, could perform better than a traditional statistical algorithm in credit card default prediction. Another approach to enhancing the performance of a fraud-related prediction on imbalanced credit data was introduced by Eteng et al. (2025) in the form of a stacked-ensemble. Even though the studies deal with more or less different prediction challenges, all of them emphasize the need to model financial risk in varied perspectives. Transactional anomalies,

customer repayment behaviour and patterns of credit use could bring to light various different facets of abnormal financial activity.

Although valuable progress has been reported in the past work, there are still several gaps. Firstly, there are numerous studies which deal with class imbalance and model design but not in a single framework. Secondly, despite the good performance of deep learning, research examining ResNet-based models in conjunction with various attention mechanisms in fraud detection is almost non-existent. Thirdly, the majority of the studies have conducted their evaluation on only one set of data, which constrains our comprehension of the generalizability of the models to other financial environments. The European Credit Card dataset in this study is a transaction-level fraud behaviour dataset with extreme class imbalance, whereas the UCI Taiwan dataset is a customer-level dataset which includes demographics information, billing history information and repayment behaviour. The two datasets thus represent two aspects of financial risk which are different yet complementary. Evaluating models across such heterogeneous datasets can provide a broader view of robustness and generalization, even if the feature spaces are not semantically identical.

The present research project is expected to fill this gap by incorporating SMOTE-ENN resampling with a ResNet-based framework with various attention mechanisms, and testing the integrated experimental framework on the European dataset, the UCI Taiwan dataset and a mixed heterogeneous environment. By doing so, the study unites three directions that have frequently been studied independently in past research: imbalance handling, attention-enhanced deep learning and evaluation on various financial datasets. The combination of the three views can be used to give a more comprehensive picture of the effectiveness of fraud detection in complex and unbalanced financial settings.

2 RESEARCH METHODS

The development of online transactions in the financial sphere triggered the growth of fraud, and sophisticated tools were necessary. Deep learning structures, especially the residual network model (ResNet), came into the limelight among several other methods because they were capable of capturing complex structures in large datasets. Nevertheless, there are still serious problems in using ResNet to detect fraud, especially in areas of data imbalance. This section outlines the full scheme of fraud detection based on integrating various sources of data, class imbalance management strategies and a deep learning model based on ResNet and using several attention mechanisms. The flowchart of the methodology of this study is displayed in Figure 1.

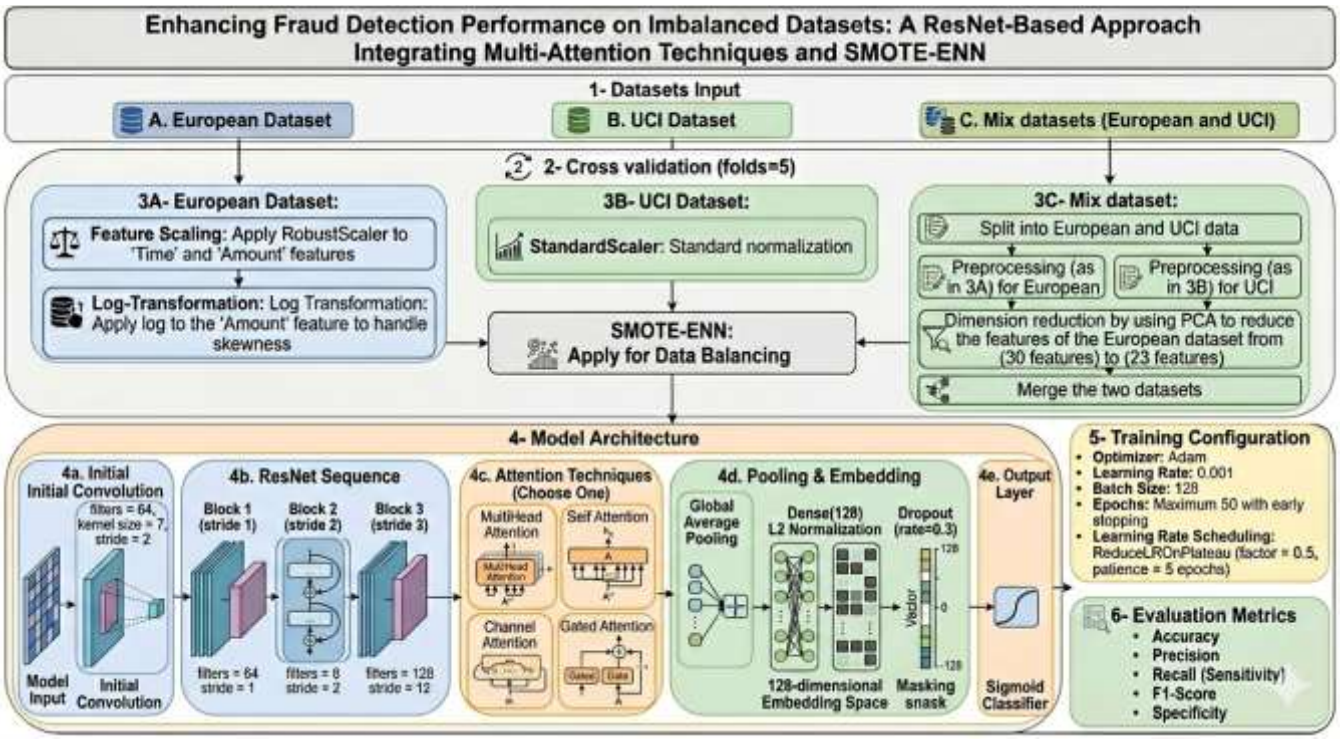

Figure 1. *Methodology of ResNet-Based approach integrating multi-attention techniques and SMOTE-ENN.*

2.1 Datasets

2.1.1 European Credit Card dataset

The European Credit Card dataset on Kaggle is a useful resource for credit card fraud detection studies. It includes the transactions of European cardholders on more than two days in September 2013 and there is a heavy imbalance in classes: the portion of fraudulent transactions is 492 out of 284,807 transactions (only about 0.172%). The majority of the features (V1-V28) are PCA-transformed due to their confidentiality, whereas the time (in seconds) and amount are not transformed. The target of the classification models is the *class* variable (0 = legitimate, 1 = fraudulent) (MLG, 2017).

2.1.2 UCI Taiwan dataset

The UCI machine learning repository has a dataset on UCI Taiwan credit card data that provides demographic variables, credit data, payment history and bill statements of credit card clients in Taiwan between April and September 2005. It consists of 30,000 observations and 25 features, one of which is the target variable (called default payment next month). These features include client demographics (e.g., age, gender, education, marital status), credit limit, payment history (past monthly payment records) and bill statement amounts. This dataset is commonly used for classification tasks, particularly to predict credit card defaults, and is valuable for developing models that assess credit risk and improve financial decision-making (Yeh, 2009).

2.1.3 Mixed dataset

Integrating the European Credit Card dataset and the UCI Taiwan dataset would result in an enriched dataset that focuses on fraud detection as well as default prediction for credit cards. The former European dataset focuses on transaction-level fraud detection using PCA features and has a severe imbalance, and the latter Taiwanese dataset provides client-level demographic information along with payment history that aids in predicting defaults. If these datasets could be fused with proper feature alignment and flexible data merging constants, it would result in many benefits. Firstly, it would enhance the depth of the analysis since it could result in building models that consider both anomalies in transactions and specific risks presented by the clients. Secondly, it would improve the reliability of the model by increasing its exposure to different data distributions as well as different feature types. Thirdly, it would aid in building sophisticated models for comprehensive risk assessment, which would enhance fraud detection as well as default prediction.

The mixed dataset was constructed as a heterogeneous financial risk classification setting rather than a semantically uniform benchmark. The European dataset represents transaction-level fraud detection, while the UCI Taiwan dataset represents client-level credit default prediction. Although these tasks differ in granularity and domain interpretation, both involve binary classification of adverse versus non-adverse financial outcomes. Accordingly, the mixed set was designed to examine whether the integrated experimental framework can maintain robust performance across different financial data distributions and risk-related behavioural patterns. Its purpose is therefore methodological, namely, to evaluate model robustness in a broader heterogeneous set, rather than to claim strict equivalence between fraud and default prediction

2.2 Cross-validation strategy

In order to have effective assessment of model performance and generalizability, especially when dealing with imbalanced data on fraud detection, 5-fold Stratified cross-validation strategy was adopted with scikit-learn Stratified Fold random seed (42) to ensure reproducibility. This approach splits the data into five folds with the original proportions of the classes fraud and non-fraud (fraud vs. non-fraud) in each of the splits, such that each of the training sets (80% of data) and test sets (20 percent) has equal proportions of both classes. Every cross-validation fold: The predictive model is fit to a new set of stratified training and not to the full set. This is then tested on the similar hold-out test data. This rigor of the procedures ensures that the resulting measures of performance will be consistent and reliable regardless of all repetitions, thereby, a good base of assessment of the predictive quality of the model.

2.3 Data preprocessing

Scaling in the case of the Fraud European data is done based on the Time and Amount features, with the assistance of Robust Scaling procedures in redressing the financial outliers in the transaction data. Also, the log-transformation of the measure's characteristic is employed to reduce the skewness of this variable's distribution, thereby improving the overall stability of the model. Fraud European dataset Preprocessing Steps:

1. Data Loading: The process began by loading the European fraud dataset.
1. Feature Scaling: RobustScaler was applied to the Time and Amount features to make them more suitable for model training while minimizing the effect of extreme values.
2. Log Transformation: Log transform the amount feature to address skewness.

The numeric variables, particularly those that include bill and payment amounts, in the UCI Credit Card Default dataset are standardised using RobustScaler, which minimises the impact of extreme values and outliers prevalent in financial data, leading to more stable and effective models. Preprocessing Steps for the UCI Credit Card dataset are:

1. Data Loading & Inspection: Load the dataset and check for missing values
2. Drop two columns: ID column and target column. The ID is simply a unique customer identifier (e.g., 1, 2, 3...). It carries no information about credit behaviour, payment history, or default risk.
3. Selective Feature Scaling: Apply StandardScaler only to specific financial amount features.

The preprocessing procedures were applied independently to each dataset before any merging operation was performed. For the European Credit Card dataset, the PCA-transformed anonymous features provided in the original dataset were retained, while the Time and Amount attributes were normalized using RobustScaler. For the UCI Taiwan dataset, the numeric variables, particularly those that include bill and payment amounts, are standardised using RobustScaler, which minimises the impact of extreme values and outliers prevalent in financial data, leading to more stable and effective models. No original features were manually removed from either dataset, except for the target labels during feature preprocessing and the ID column from the UCI dataset.

Since the two datasets differed substantially in feature structure and dimensionality, dimensional harmonization was required before constructing the mixed dataset. Principal Component Analysis (PCA) was applied to reduce the feature space of the European dataset (has 30 features) to a dimensionality compatible with the UCI Taiwan dataset (has 23 features). The purpose of this transformation was not to establish semantic equivalence between features, but rather to provide a consistent numerical representation suitable for joint model training and comparative robustness evaluation. The selected dimensionality was chosen to preserve the majority of variance while maintaining compatibility between the heterogeneous datasets. Accordingly, the mixed dataset should be interpreted as a heterogeneous exploratory evaluation set rather than as a semantically unified financial benchmark.

2.4 Class imbalance handling

The imbalance in classes is one of the biggest obstacles in machine learning, particularly in sensitive areas such as healthcare and finance, where it can have a significant impact on the classification accuracy. A combination of SMOTE and ENN can be considered one of the effective solutions to the issue, as it addresses it in two directions: it synthesizes the samples of the minority group and removes the noisy data cases. The applicability of this method in enhancing the performance of classification has been noted in recent studies. Moreover, Tram and Bichanga (2025) showed that the combination of SMOTE-ENN and deep learning models has a great impact on improving the performance in credit card fraud detection (Bonde & Bichanga, 2025; Hairani & Priyanto, 2023; Nishat et al., 2022).

2.5 Model architecture

An adapted ResNet model, when combined with attention systems, significantly extends the feature learning as well as the pattern recognition capabilities. The ResNet architecture is one-dimensional and it begins with an initial convolutional layer (64 filters, kernel size 7, stride 2), followed by a ReLU nonlinearity and batch normalization. Following layers are a series of residual blocks which increasingly enlarge the network; that is, two blocks use 64-filter convolution and the subsequent two blocks use 128-filter convolution. Each of the remaining blocks has two

convolutional layers, each having batch normalization and ReLU activations, a skip connection and strides which are adjusted incrementally.

The attention layer is directly followed by the ResNet blocks. This layer was added because the model should focus on the most informative features that are obtained by the residual blocks. Although ResNet blocks can effectively learn deep hierarchical features, they assign the same level of significance to all the features. This may quench the significance of the critical pattern especially in situations where the critical pattern is faced with obscured and intricate datasets, including those in the investigation of financial dealings. A subsequent addition of the attention layer allows the network to learn to attend to the most salient signals in the dense set of feature maps generated by ResNet and thus learn to selectively focus on important relationships and ignore irrelevant noise. The resulting architecture allows the network to maintain the strong feature extraction properties of ResNet and add to it contextual information and feature prioritization, which end up producing predictions that are more accurate and interpretable.

We compare four different attention mechanisms, namely self-attention, channel attention, gated attention units and multihead attention. The self-attention block is an auxiliary element of a ResNet-based architecture, which allows the network to move past local pattern recognition and establish global interconnections that exist among the input data. It is possible for ResNet neural procedures to overlook long-range dependencies that stretch to broader geographical or temporal scales, although they accomplish a high level of success in recognising short-range, localised characteristics, such as spikes in transaction values. The model has a self-attention module that makes it suited to evaluate the mutual relevance of each feature in a sequence or feature map against each other feature, thus creating a context-dependent representation. The ResNet is therefore advantaged by a combination of its natural ability to extract local features and an enhanced global awareness making it more proficient in identifying the presence of delicate and broadly spread patterns, such as fraudulent patterns, which are distributed among many transactions. Research has shown that integrating attention into ResNet architectures improves performance in sequence modelling and computer vision by capturing dependencies that convolution alone cannot (Vaswani et al., 2017; Wang et al., 2018).

Channel attention is a mechanism that teaches the network to focus on the most important feature channels while suppressing less useful ones. In ResNet, each convolutional layer produces many feature maps (channels), but not all of them contribute equally to the task at hand. Channel attention works like a "spotlight": it analyses all channels, determines which ones are most informative (for example, those strongly signalling fraud or a specific object pattern) and then reweights them so that the model pays more attention to the right features. Technically, this is often implemented with a squeeze-and-excitation process, where global information across each channel is pooled, passed through small dense layers and used to rescale the original features. The model is more selective and efficient due to this approach as it concentrates on important features and disregards noise. It enables these networks to be more accurate with the channels that are important, at minimal computational costs (Hu et al., 2018).

The gated attention unit (GAU), which causes the selective transmission of one feature map property and the suppression of others, plays a similar role in the ResNet architecture. Though there are some residual links between the exemplar and encoder which lead to the development of more salient and rich encoding representations, it is important to mention that not all of the extracted features are equally predictive of the downstream task; in fact, some of them can as well be associated with harmful noise. To control the information flow in an intelligent way, the GAU uses a gating mechanism, often a differentiable sigmoid function. In this process, the scalar weights are confined within the range of [0, 1] and in the latter case, peripheral or subsidiary feature activations are assigned smaller or null weights and the salient feature is assigned larger weights, so that it can propagate without being weakened significantly. These kinds of dynamic suppression improve concentration and network efficiency (Dauphin et al., 2017).

Multihead attention applied in a ResNet architecture offers the network the capacity to examine input data on numerous related orientations simultaneously, which considerably boosts its capabilities of understanding complicated feature correlations. Unlike a single self-attention layer, which is only capable of modelling a single kind of relationship between the features, a multihead system divides the feature space into multiple different heads. The two heads are also charged with the responsibility of learning different patterns or dependencies; e.g., one head can focus on short-term correlations and another can focus on long-term trends. After calculating a specific attention map of the model through each of the heads, the resulting representations are then fused, giving the model a highly

insightful, multi-view representation of the information. By incorporating this mechanism into ResNet, it becomes a secondary component and a complement to the original proficiency of the convolutional layers in extracting local structure, providing instead a global, multi-scale sensitivity that cuts across channels as well as spatial or time scales (Vaswani et al., 2017; Wang et al., 2018). In other words, the method gives a ResNet the ability to analyse the same data with multiple aspects of analysis, each revealing a different facet of the entire construct.

After the attention mechanism, there are a number of other elements that are added to the architecture in the next stage. It consists of a global average pooling operation, a flattening layer, a fully connected dense layer, a dropout regularizer and finally an output layer. These layers have a given application in refining the representations learned as well as the model prediction consistency. The training procedure is coordinated per fold, with a particular setup being used per iteration. Such a systematic method ensures that the model is tested on a whole dataset and, therefore, a strong evaluation of the model is made. The binary cross-entropy is employed as the loss function. A batch size of 128 samples per batch is utilised. Figure 1 illustrates the architecture of the integrated experimental framework. The pseudocode of the methodology is shown in Algorithm 1.

Algorithm 1. *Methodology pseudocode.*

```
Input: Three datasets (European, UCI Taiwan, mixed)
Output: The results of metrics used for each dataset.
1:          Load dataset
2:          Preprocess data
            - Scale 'time' and 'amount'
            - Apply log-transform on 'amount'

3:          Define model
            - Build 1D ResNet with residual blocks
            -  Add  attention  layer  (multihead/self/channel/gated
            attention units)
            - Global pooling + dense layers
            - Output: sigmoid (binary classification)
4:          Cross-validation training
            FOR each fold in StratifiedKFold:
            - Split into train/test
            - Balance data using SMOTEENN
            - Compile model (Adam, binary crossentropy, accuracy)
            - Train with early stopping + checkpoint
            - Load best model
5:          Evaluation
            - Predict on test data
            - Calculate metrics:
            Accuracy, precision, recall, F1, AUC
            - Save metrics to CSV
            END
```

2.6 Implementation details

All the experiments were conducted using 5-fold stratified cross-validation with shuffle=True and random_state=42. SMOTE-ENN was instantiated with random_state=42. The attention modules were configured as follows: the self-attention layer used 128 units for the query, key and value projections; the channel attention module used a reduction ratio of 16; the gated attention unit used a number of units equal to the input feature channel dimension; and the multihead attention module used 8 attention heads with a key dimension of 32. A dropout with a rate of 0.3 was applied in the classifier head before the final sigmoid output layer. Both early stopping and model checkpointing were based on validation AUC (val_auc). Training was performed for up to 50 epochs with a batch size of 128. Early stopping used a patience of 15 epochs with restoration of the best weights, while checkpointing saved the best-performing model in max mode. In addition, the learning rate was adjusted dynamically using ReduceLROnPlateau with factor 0.5, patience 1, cooldown 1 and a minimum learning rate of 1e-7. The training duration for each dataset

varied; the minimum was for the UCI dataset, at about 9.1517045 minutes. The experiments were conducted on a computer with an Intel Core i7 processor, 40 GB RAM and an NVIDIA RTX 2070 GPU with ram 8 GB running Windows 11.

3 EXPERIMENT DESIGN

This study consists of an experimental design that has three subsections as discussed here: (i) the evaluation metrics of models, (ii) presentation of comparative studies and (iii) an analysis of Kendall's W statistic that can be used to rank the integrated experimental framework of previous studies.

3.1 Model evaluation metrics

Metrics for evaluating the integrated experimental framework are crucial to assessing its performance. In this study, the following metrics are used to evaluate the model:

- Accuracy shows how often the model gets its predictions right overall. It counts all the times it guessed correctly—both "yes" and "no"—and compares that to all the guesses that it made. However, it is not useful when dealing with unbalanced data (Mandal & Mahto, 2023).
- Precision is concerned with the quality of the "yes" guesses of the model. It tells us, among the occasions when the model said "no", how many times it was wrong. This plays a crucial role when we do not want false alarms whereby the model answers "yes", yet it should not.
- Recall considers things differently: it is a measurement of the number of the actual cases of "yes" that the model was able to find. This is important when it is a big issue to miss a real "yes" and we desire to snatch as many as we can.
- The F1 score balances precision and recall. It is a way to make sure that the model is not merely guessing "yes" carefully or trying to catch every "yes" at the expense of mistakes. It gives us a single score that rewards a good balance between both.
- AUC (area under the ROC curve) assesses the ability of a model to distinguish between positive and negative classes in different threshold settings. By plotting the true positive rate against the false positive rate, the AUC represents the area under this curve, with higher values indicating better model performance, and is less sensitive to class imbalance than accuracy.

3.2 Comparison studies

The second strategy to evaluate the integrated experimental framework is to compare the outcomes with the results of a comparative benchmarking of existing approaches that use the same datasets and metrics, which are listed below.

3.2.1 Studies working with European Credit Card dataset

Alarfaj et al. (2022) discussed machine learning and deep learning to identify credit card fraud. On the European dataset, they tested various algorithms, such as CNNs, and tried to adjust the model settings to achieve improved results, such as the number of layers in the model and its training duration. Jiang et al. (2023) presented a new system, named UAAD-FDNet. It views fraud as odd or anomalous behaviour and trains autoencoders with attention and GANs to distinguish between fraud and regular transactions with no labelled data. Zhao and Guan (2023) introduced a method called CTCN, which uses a special GAN to generate extra data and balance the dataset, along with a temporal convolutional network to pick up on time-related patterns. They also cleaned up confusing data points to help the model make better decisions.

3.2.2 Studies working with UCI Taiwan dataset

The work of Zhao and Guan (2023) was explained in the previous subsection. Emmanuel et al. (2024) proposed a machine learning-based credit risk prediction system using a stacked classifier combined with a filter-based feature selection (FS) method. Bhandary and Ghosh (2025) found that modern machine learning techniques, particularly deep neural networks (DNN), significantly outperformed traditional statistical methods regarding prediction accuracy, F1 score, G-mean and area under the ROC curve (AUC).

The data are skewed since there are very few cases of fraud as opposed to normal transactions and the use of traditional accuracy is therefore not very reliable. Rather, metrics such as AUC (area under the ROC curve) and the F1 score would be more appropriate in comparing the methodology with the existing work on imbalanced data. AUC is not as sensitive to the class imbalance as plain accuracy is because it can detect frauds and valid transactions at various decision thresholds. In the meantime, the F1 score provides a reasonable balance of precision and recall and provides a reasonable picture of the model performance, which is of particular importance in fraud detection, where we do not want to either get a false alarm or miss a fraud event.

3.3 Kendall's coefficient of concordance

The third important statistic that can be implemented to compare the methods offered by previous research to rank them is Kendall's coefficient of concordance. It is a non-parametric statistic indicating the degree of agreement between several raters ranking a set of items. In essence, it measures the degree to which the rankings are consistent so that a value of agreement 0 would mean that there exists no agreement and a value of agreement 1 would mean the highest degree of agreement. The statistic is especially applicable in circumstances where a subjective judgment is at stake, e.g., where the expertise of opinions is considered, inter-rater reliability or rankings can be analysed in different areas (Altalib & Salim, 2022).

In the present study, Kendall's coefficient of concordance is used as a descriptive measure of agreement between the rankings produced by the two selected evaluation metrics, namely F1 score and AUC. Its purpose is not to replace these standard performance measures, but to determine whether they yield a consistent relative ordering of the competing models. In this way, Kendall's W provides complementary insight into the stability of comparative conclusions across imbalance-sensitive criteria, whereas F1 score and AUC remain the primary measures of predictive performance.

4 RESULTS AND DISCUSSION

As shown in the tables, the experimental results indicate how four various ResNet-based attention mechanisms (ResNet multihead attention, ResNet self-attention, ResNet channel attention and ResNet gated attention) perform following the application of the SMOTE-ENN resampling technique that is generally used when dealing with imbalanced data. The performance is compared over three different datasets namely: the European dataset, the UCI dataset and a mixed dataset of European and UCI data.

Table 1. *Mean and standard deviation (SD) of experimental results of four different ResNet-based attention mechanisms on European dataset.*

	Integrated experimental frameworks	Accuracy	Precision	Recall	F1 score	AUC
Mean	ResNet multihead attention	99.91	85.45	92.72	88.71	97.30
	ResNet self-attention	99.90	83.83	92.82	87.79	97.25
	ResNet channel attention	99.91	85.19	92.72	88.56	97.78
	ResNet gated attention	99.91	85.22	92.01	88.27	97.80
SD	ResNet multihead attention	0.000141845	0.023526	0.004233	0.013686	0.006361
	ResNet self-attention	0.0000654	0.010024	0.0051	0.006734	0.002496
	ResNet channel attention	0.000112827	0.019232	0.00659	0.009893	0.001258
	ResNet gated attention	0.000117473	0.021303	0.007673	0.010347	0.00383

Table 2. *Mean and standard deviation (SD) of experimental results of four different ResNet-based attention mechanisms on UCI dataset*

	Integrated experimental frameworks	Accuracy	Precision	Recall	F1 score	AUC
Mean	ResNet multihead attention	75.42	67.15	71.01	68.24	77.25
	ResNet self-attention	73.14	65.59	69.05	66.10	74.75

	Integrated experimental frameworks	Accuracy	Precision	Recall	F1 score	AUC
	ResNet channel attention	74.61	66.74	71.01	67.78	77.23
	ResNet gated attention	76.32	67.89	71.28	68.93	77.42
SD	ResNet multihead attention	0.004514599	0.007075	0.004316	0.0058	0.030598
	ResNet self-attention	0.036064028	0.034448	0.048204	0.048935	0.036548
	ResNet channel attention	0.007021896	0.003111	0.008626	0.006457	0.025732
	ResNet gated attention	0.012803828	0.004168	0.012924	0.006883	0.027875

Table 3. *Mean and standard deviation (SD) of experimental results of four different ResNet-based attention mechanisms on mixed dataset.*

	Integrated experimental frameworks	Accuracy	Precision	Recall	F1 score	AUC
Mean	ResNet multihead attention	91.05	59.48	89.01	63.39	89.87
	ResNet self-attention	92.25	61.22	95.41	66.24	97.40
	ResNet channel attention	92.37	61.36	95.56	66.46	98.00
	ResNet gated attention	92.22	61.20	95.51	66.20	98.00
SD	ResNet multihead attention	0.037862057	0.139051	0.062365	0.176133	0.256731
	ResNet self-attention	0.004364428	0.003332	0.00695	0.009785	0.002911
	ResNet channel attention	0.002289734	0.002126	0.003581	0.001226	0.002789
	ResNet gated attention	0.003216057	0.002251	0.005084	0.000916	0.002502

The experimental results reveal consistent performance patterns across the three datasets that reflect their inherent characteristics and the suitability of different attention mechanisms. On the European dataset, all four integrated experimental frameworks are almost perfect (~99.9%) in terms of accuracy, with AUC scores of more than 97, meaning that the classes are highly separable and the features are cleanly distributed. The best here is ResNet-multihead attention (F1: 88.71) since parallel attention heads can distinguish various types of frauds in case the data are regular and not noisy. The small standard deviations of all the measures are indicative of the outstanding stability of the data.

The UCI dataset is in a sharp contrast with an accuracy of about 75% and AUC of about 77 as it is inherently complex, has correlated features and overlapping distributions of classes. ResNet gated attention achieved the highest accuracy (76.3) and F1 score (68.9) among the evaluated configurations. One possible explanation is that the gating mechanism may help suppress less informative feature representations. The instability of self-attention (SD: 0.036) is considerably larger than that of the other mechanisms, indicating that it fits well to particular patterns of noise without the regularization effects of gating or multihead diversity and that the adaptive feature weighting of gated attention is important to noisy credit data in the real world.

The results of the mixed dataset show intermediate results (~92% accuracy) more similar to the European results, which suggests that the cleaner European data probably dominate or level off the combined distribution. However, an extreme precision-recall imbalance emerges (precision ~61%, recall ~95%), where frameworks aggressively predict defaults due to conflicting patterns between datasets and aggressive SMOTE-ENN resampling. ResNet channel attention excels here (accuracy 92.4%, AUC 98.0%) because its adaptive channel recalibration handles the domain shift between the European and UCI distributions by learning which features remain reliable across both sources. Notably, multihead attention produced lower average performance and greater variability than the other evaluated configurations in the mixed-dataset.

Table 4. *Comparisons between integrated experimental frameworks and comparative benchmarking studies using European dataset.*

Previous studies	F1 score	AUC
Alarfaj et al. (2022)	85.71	98.00
Zhao & Guan, 2023	81.87	91.47
Jiang et al. (2023)	84.89	94.37
Jiang et al. (2023)	85.29	95.15
Zhao et al. (2023)	81.87	91.47
ResNet multihead attention	88.71	97.30
ResNet self-attention	87.79	97.25
ResNet channel attention	88.56	97.78
ResNet gated attention	88.27	97.80

Table 5. *Comparisons between integrated experimental frameworks with comparative benchmarking studies based on UCI dataset.*

Previous studies	F1 score	AUC
Bhandary et al. (2025)	47.32	77
Bhandary et al. (2025)	48.2	77
Zhao et al. (2023)	53.49	84.69
Emmanuel et al. (2024)	51.35	87
ResNet multihead attention	68.24	77.25
ResNet self-attention	66.10	74.75
ResNet channel attention	67.78	77.23
ResNet gated attention	68.93	77.42

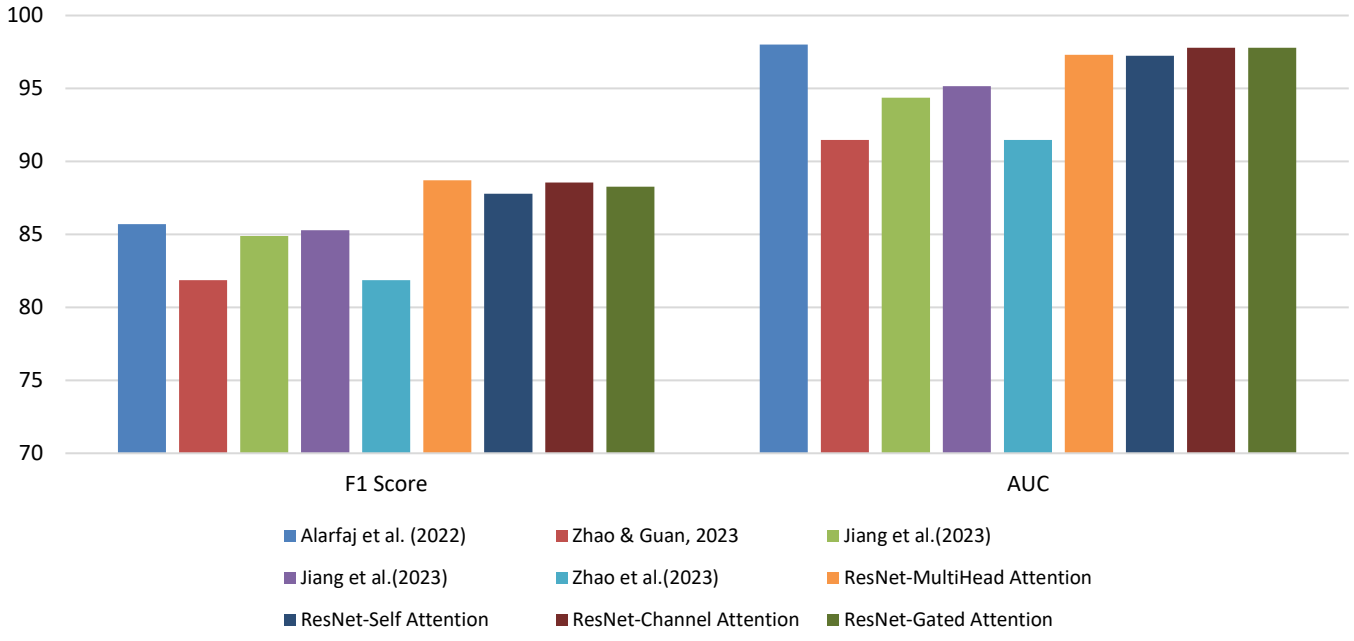

Figure 2. *Comparison of integrated experimental frameworks with comparative benchmarking studies founded on European dataset.*

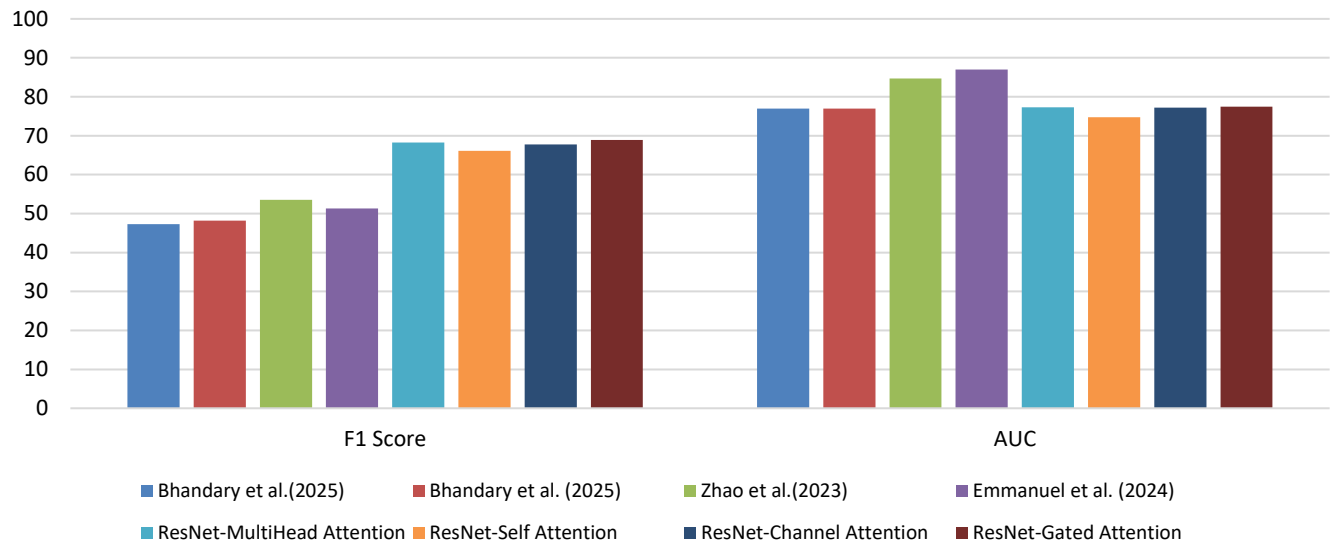

Figure 3. *Comparison of integrated experimental frameworks with comparative benchmarking studies founded on UCI dataset.*

Furthermore, the second plan was to compare the integrated experimental frameworks using the methods that utilise the same datasets and measures. The results presented in Tables 4 and 5 indicate that the evaluated ResNet attention configurations achieved competitive performance relative to the values reported in the selected benchmarking studies. On the European dataset, the evaluated configurations produced F1 scores ranging from 87.79% to 88.71%, which are higher than the F1 scores reported in the selected comparison studies. However, their AUC values were not consistently higher than all previously reported results. On the UCI Taiwan dataset, the evaluated configurations similarly achieved higher F1 scores than the selected reported values, while several previous studies reported higher AUC values. These results suggest that the present configurations provide a favourable balance between precision and recall within the present experimental setting, but they should not be interpreted as evidence of universal superiority since the compared studies may differ in preprocessing, data partitioning, resampling, threshold selection and validation procedures.

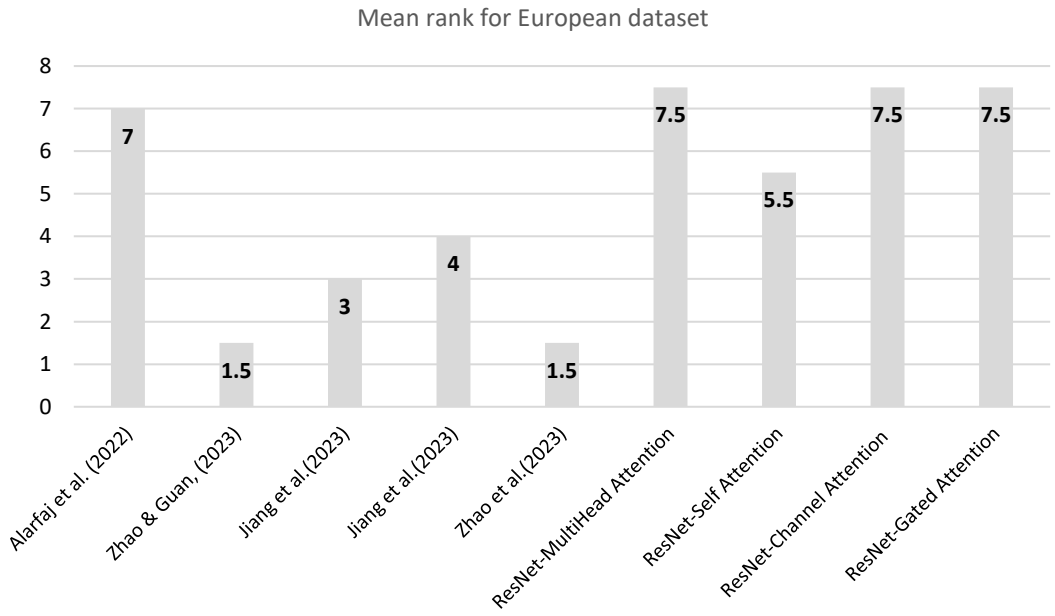

Figure 4. *Ranking of all methods based on European dataset using Kendall's W.*

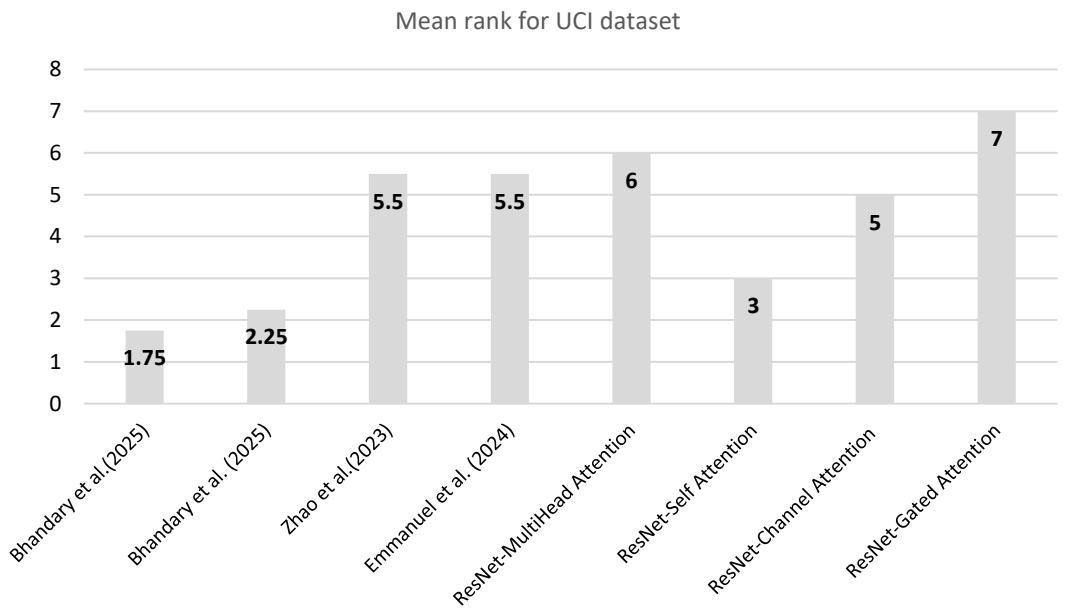

Figure 5. *Ranking of all methods based on UCI dataset using Kendall's W.*

Moreover, Figures 4 and 5 show the general ranking of all the methods using Kendall's W, indicating that some of the integrated experimental frameworks have a comparative evaluation of established methods and have the top rank in all the datasets used.

The mean rank findings are the relative rankings of the compared methods in terms of F1 score and AUC. The increase in the mean rank values is a sign of improved general status within the two evaluation criteria. Figure 4 indicates that the integrated experimental frameworks demonstrated the highest average rankings on the European dataset. Specifically, the ResNet-multihead attention, ResNet channel attention and ResNet gated attention variants all obtained the highest mean rank of 7.50, which is the best and most reliable performance. The ResNet self-attention type was followed by an average rank of 5.50. Alarfaj et al. (2022) scored 7.00 in the list of the previous studies and the methods of Jiang et al. (2023), as well as Zhao and Guan (2023) and Zhao et al. (2023), were ranked among the lowest mean rankers (1.50). These results show that the offered frameworks, especially the multihead, channel and gated attention designs, offer a positive overall ranking of the chosen performance indicators.

The integrated experimental frameworks scored highly on average, with the ResNet gated attention framework achieving the best overall score of 7.00, which means the most favourable and consistent performance of all the methods compared on the UCI dataset in Figure 5. The ResNet multihead attention framework had a mean rank of 6.00 and both Zhao et al. (2023) and Emmanuel et al. (2024) had a mean rank of 5.50, which is competitive. The ResNet channel attention framework also performed well with a mean rank of 5.00. In contrast, the ResNet self-attention framework obtained a lower mean rank of 3.00, indicating weaker relative performance compared with the other integrated experimental frameworks. The lowest mean ranks was observed for Bhandary et al. (2025) with values of 1.75 and 2.25. Overall, these results suggest that the integrated experimental frameworks, particularly the ResNet gated attention, achieved a stable and favourable ranking when evaluated jointly using F1 score and AUC for the UCI dataset.

4.1 Ablation study

To better understand the contribution of the components of the integrated experimental frameworks, an ablation study was conducted by removing the attention mechanism and the SMOTE-ENN balancing strategy while keeping all other model settings unchanged. The analysis was performed on the European, UCI and mixed datasets using accuracy, precision, recall, F1 score and AUC as evaluation metrics, as shown in Table 6.

Table 6. *Ablation study results showing impact of attention mechanisms and SMOTE-ENN on fraud detection performance across European, UCI and mixed datasets.*

Datasets	Type of framework	Accuracy	Precision	Recall	F1 score	AUC
European dataset	Full framework	99.91	85.22	92.01	88.27	97.8
	Without attention	97.80	54.65	94.84	57.62	97.68
	Without SMOTE-ENN	97.28	53.04	93.06	54.96	95.96
UCI dataset	Full framework	76.32	67.89	71.28	68.93	77.42
	Without attention	66.46	63.12	68.25	61.79	75.32
	Without SMOTE-ENN	75.78	67.31	70.43	68.16	75.37
Mixed dataset	Full framework	92.22	61.20	95.51	66.20	98.00
	Without attention	92.20	61.19	95.39	66.20	97.90
	Without SMOTE-ENN	90.10	59.40	93.21	63.06	96.85

For the European dataset, the complete framework achieved the best overall performance, obtaining an F1 score of 88.27% and an AUC of 97.8%. When the attention mechanism was removed, the F1 score dropped substantially to 57.62%, despite a slight increase in recall. This result indicates that the attention module plays a crucial role in improving the ability of the model to distinguish fraudulent transactions from legitimate ones. In particular, the large decrease in precision suggests that the framework generated considerably more false alarms when attention was absent. Similarly, removing SMOTE-ENN resulted in a further reduction in F1 score to 54.96% and a decrease in AUC to 95.96%. These findings demonstrate that the balancing strategy contributes significantly to learning representative patterns from the minority fraud class and improving overall classification quality.

A similar trend can be observed in the UCI dataset. The complete framework achieved the highest values across all the evaluation metrics, with an F1 score of 68.93% and an AUC of 77.42%. Excluding the attention mechanism reduced the F1 score by more than seven percentage points and lowered the AUC, indicating that attention improves feature representation and helps the framework focus on the most informative transaction characteristics. When SMOTE-ENN was removed, performance also decreased, although the reduction was less pronounced than that observed when attention was excluded. This suggests that both components contribute positively to the framework, with the attention mechanism having a slightly greater influence on the UCI dataset.

The results obtained from the mixed dataset provide additional insights into the behaviour of the integrated experimental framework when trained on heterogeneous data sources. The complete framework achieved the highest overall performance, with an AUC of 98% and a recall of 95.51%. Removing the attention mechanism had only a marginal effect on the performance metrics, indicating that the mixed dataset already contains sufficiently rich and diverse information for the framework to learn effective representations. Nevertheless, the full framework still achieved the highest AUC value, suggesting that attention contributes to more reliable discrimination between fraud and non-fraud classes. In contrast, removing SMOTE-ENN led to a noticeable decline in all the evaluation metrics, particularly recall, F1 score and AUC. This behaviour highlights the importance of addressing class imbalance when dealing with combined datasets containing transactions from multiple sources.

Overall, the ablation study confirms the effectiveness of both the attention mechanism and the SMOTE-ENN balancing strategy. The attention component improves the model ability to focus on relevant fraud-related features and reduces the occurrence of false positive predictions, leading to substantial gains in precision and F1 score. Meanwhile, SMOTE-ENN enhances minority class representation and helps the model learn more discriminative fraud patterns, resulting in improved recall and AUC values. These results support the usefulness of integrating the two components in the present pipeline, but they should not be interpreted as evidence of universal superiority across all financial datasets, nor does the integrated experimental framework constitute a novel methodological architecture.

5 CONCLUSION

Our approach is a combination of the ResNet models and a variety of attention mechanisms with the intelligent data balancing mechanism called SMOTE-ENN, on both the European Credit Card and UCI Taiwan datasets and a

combination of both datasets. The experimental results demonstrate that the integrated ResNet-based attention framework achieved competitive performance across the evaluated financial risk datasets, particularly in terms of F1 score under highly imbalanced conditions.

However, although the integrated experimental frameworks performed favourably overall, the AUC is not consistently superior across all settings and the most consistent gains are observed in F1 scores or ranking consistency. Accordingly, the findings should be interpreted as evidence of competitive empirical effectiveness rather than universal superiority across all evaluation metrics. In the future, this method could be expanded to other financial datasets and fine-tuned further to make it even more flexible and accurate.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: M.K.A.: Conceptualization, Methodology, Software, Validation, Writing – Original draft preparation, Writing – Reviewing and Editing. K.A.: Data curation, Visualization, Investigation. N.S.: Supervision.

Statement on the Use of Artificial Intelligence Tools: During manuscript revision, the authors used ChatGPT, developed by OpenAI, to refine language and improve clarity in some selected sections of the manuscript. The tool was not used to generate research data, conduct analysis, make methodological decisions, or replace authorial interpretation. All AI-assisted outputs were critically reviewed, edited, and approved by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Data Availability: European Credit Card Fraud Detection Dataset is freely available on the Kaggle site: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. UCI Credit Card Default Dataset is freely available on the site: https://doi.org/10.24432/C55S3H.

REFERENCES

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, 10, 39700–39715. https://doi.org/10.1109/ACCESS.2022.3166891

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, 12(19), 9637. https://doi.org/10.3390/app12199637

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *IEEE Access*, 11, 137188–137203. https://doi.org/10.1109/ACCESS.2023.3339226

Altalib, M. K., & Salim, N. (2022). Hybrid-Enhanced Siamese Similarity Models in Ligand-Based Virtual Screen. *Biomolecules*, 12(11), 1719. https://doi.org/10.3390/biom12111719

Azhar, N. A., Pozi, M. S. M., Din, A. M., & Jatowt, A. (2022). An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis. *IEEE Transactions on Knowledge and Data Engineering*, 35(7), 6651–6672. https://doi.org/10.1109/TKDE.2022.3179381

Bhandary, R., & Ghosh, B. K. (2025). Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods. *Journal of Risk and Financial Management*, 18(1), 23. https://doi.org/10.3390/jrfm18010023

Bonde, L., & Bichanga, A. K. (2025). Improving Credit Card Fraud Detection with Ensemble Deep Learning-Based Models: A Hybrid Approach Using SMOTE-ENN. *Journal of Computing Theories and Applications*, 2(3), 384. https://doi.org/10.62411/jcta.12021

Bounab, R., Zarour, K., Guelib, B., & Khlifa, N. (2024). Enhancing Medicare fraud detection through Machine Learning: Addressing class imbalance with SMOTE-ENN. *IEEE Access*, 12, 54399–54413. https://doi.org/10.1109/access.2024.3385781

Cao, R., Liu, G., Xie, Y., & Jiang, C. (2021). Two-Level Attention Model of Representation Learning for Fraud Detection. *IEEE Transactions on Computational Social Systems*, 8(6), 1291–1301. https://doi.org/10.1109/tcss.2021.3074175

Chen, Y., Zhao, C., Xu, Y., & Nie, C. (2025). Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. *arXiv* preprint arXiv:2502.00201. https://doi.org/10.48550/arXiv.2502.00201

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). Language modeling with gated convolutional networks. In *Proceedings of the 34th International Conference on Machine Learning*, (pp. 933–941). JMLR.

Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data*, 11(1), 23. https://doi.org/10.1186/s40537-024-00882-0

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, 10, 16400–16407. https://doi.org/10.1109/access.2022.3148298

Eteng, I. E., Chinedu, U. L., & Ibor, A. E. (2025). A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets. *Journal of the Nigerian Society of Physical Sciences*, 7(1), 2066. https://doi.org/10.46481/jnsps.2025.2066

Hairani, H., & Priyanto, D. (2023). A new approach of hybrid sampling SMOTE and ENN to the accuracy of machine learning methods on unbalanced diabetes disease data. *International Journal of Advanced Computer Science and Applications*, 14(8), 585–590. https://doi.org/10.14569/IJACSA.2023.0140864

Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 7132–7141). IEEE. https://doi.org/10.1109/CVPR.2018.00745

Huang, H., Wang, P., Pei, J., Wang, J., Alexanian, S., & Niyato, D. (2025). Deep Learning Advancements in Anomaly Detection: A Comprehensive survey. *IEEE Internet of Things Journal*, 12(21), 44318–44342. https://doi.org/10.1109/jiot.2025.3585884

Ileberi, E., & Sun, Y. (2024). A hybrid deep learning ensemble model for credit card fraud detection. *IEEE Access*, 12, 175829–175838. https://doi.org/10.1109/access.2024.3502542

Jiang, S., Dong, R., Wang, J., & Xia, M. (2023). Credit card fraud detection based on unsupervised attentional anomaly detection network. *Systems*, 11(6), 305. https://doi.org/10.3390/systems11060305

Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, 8(1), 6. https://doi.org/10.3390/bdcc8010006

Lindemulder, G. (2024). What is fraud detection. IBM. https://www.ibm.com/think/topics/fraud-detection

Mandal, P. K., & Mahto, R. V. (2023). Deep multi-branch CNN architecture for early Alzheimer's detection from brain MRIs. *Sensors*, 23(19), 8192. https://doi.org/10.3390/s23198192

Mienye, I. D., & Jere, N. (2024). Deep Learning for Credit Card Fraud Detection: A review of Algorithms, challenges, and solutions. *IEEE Access*, 12, 96893–96910. https://doi.org/10.1109/access.2024.3426955

Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access*, 11, 30628–30638. https://doi.org/10.1109/access.2023.3262020

MLG. (2017). Credit Card Fraud Detection. *Kaggle*. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Nishat, M. M., Faisal, F., Ratul, I. J., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., Reza, M. T., & Khan, M. R. H. (2022). A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. *Scientific Programming*, 2022, 1–17. https://doi.org/10.1155/2022/3649406

Nguyen, T., Khadka, R., Phan, N., Yazidi, A., Halvorsen, P., & Riegler, M. A. (2023). Combining datasets to improve model fitting. In *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE. https://doi.org/10.1109/IJCNN54540.2023.10191273

Saeed, A. Q., Aldulaimi, M. H., Ismail, I. A., Ahmed, I. M., Yahya, Y. A., Kharma, Q. M., & Ghazal, T. M. (2024). Integrating Three Machine Learning Algorithms in Ensemble Learning Model for Improving Content-based Spam Email Recognition. *Journal of Soft Computing and Data Mining*, 5(2), 188–196.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*. NIPS.

Wang, X., Ross, G., Gupta, A., & He, K. (2018). Non-Local Neural Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/CVPR.2018.00813

Yeh, I. (2009). Default of Credit Card Clients [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C55S3H

Younis, M. C., Fadil, Z. J., & Bahnam, B. S. (2025). IoT-Based System for Human Localization Activity Recognition Using Hybrid Deep Learning Techniques. *Inteligencia Artificial*, 28(75), 298–314. https://doi.org/10.4114/intartif.vol28iss75pp298-314

Zhao, X., & Guan, S. (2023). CTCN: a novel credit card fraud detection method based on Conditional Tabular Generative Adversarial Networks and Temporal Convolutional Network. *PeerJ Computer Science*, 9, e1634. https://doi.org/10.7717/peerj-cs.1634

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951