Journal of Science and Technology ISSN: 2362-1443
2023, Vol. 3, No. 1, 31-39

Credit Card Fraud Detection using Heterogeneous Ensemble Techniques

Bhim Prasad Ale
Gandaki College of Engineering and Science
msc2018ise9@gces.edu.np

Ashim Khadka
Nepal College of Information Technology
ashim.khadka@ncit.edu.np

***Article History*:**
Received: 10 July 2023
Revised: 7 October 2023
Accepted: 26 November 2023

Keywords—Credit Card Fraud Detection, Ensemble Techniques, Feature Selection, Heterogeneous, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and Support Vector Machine (SVM).

***Abstract*—Credit card fraud is a major issue in the financial sector, necessitating advanced detection methods to minimize financial losses and improve security. Traditional techniques often struggle with the high dimensionality of transaction data and the imbalanced nature of fraud datasets. To effectively detect fraudulent transactions, this paper proposes a heterogeneous ensemble learning approach that combines the best selected k features and voting techniques with SVM, ANN, and k-NN classifiers. Heterogeneous ensembles combine classifiers that utilize different algorithms, representations, or even feature sets. This diversity allows the ensemble to capture a wider range of patterns and relationships in the unseen data. The proposed ensemble model outperforms the individual classifiers, demonstrating the superior performance of the ensemble in mitigating challenges associated with credit card fraud detection. The ensemble performs better than existing models in terms of accuracy, precision, recall, and $F_{0.1}$ score metrics. The recall rate, which is crucial for spotting fraudulent transactions, is improved by the ensemble technique, which effectively balances the trade-offs between precision and recall. The results indicate that feature selection using k-best significantly enhances the performance of individual classifiers and soft-voting ensemble methods. The findings presented lay the groundwork for future advancements in the development of more robust and adaptive fraud detection systems. The improved performance shows that the proposed approach is effective in detecting credit card fraud, and it has the potential to be implemented in the real world.**

I. INTRODUCTION

Credit card allows cardholders to make purchases and pay for services based on credit extended to them by the card issuer (usually a bank). It's a convenient and widely accepted form of payment that allows transactions to be completed electronically, typically with the promise of repayment to the issuer at a later date. With the rise of the internet, cash transactions have shifted to cashless with the use of credit cards in e-commerce, point of sale, tap and pay, and many online billing systems. Therefore, many fraudsters are looking for opportunities to exploit online payment. To minimize the risk and increase security, the magnetic stripe has been replaced with Europay, Mastercard, and Visa (EMV) chip data, tokenization in mobile pay like Google Pay, Apple Pay, Samsung Pay, and 3DS secure for e-commerce transactions [1], [2]. However, even with such measures, fraud in credit card transactions is not fully protected. Around 1.5 billion fraud transactions occur annually in the European market [3], [4]. Financial Fraud Action (FFA) UK reports fraud losses on UK-issued cards totalled £574.2 million in 2020, a seven percent fall from £620.6 million in 2019. Fig. 1 shows that most of the frauds are remote purchases, carried out online via the Internet, where the card is not present.

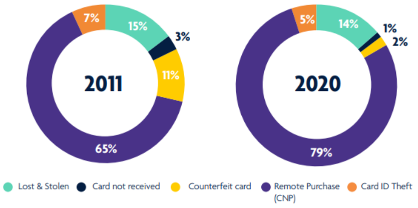

Fig. 1. Losses by type of Fraud, as a % of total loss [5]

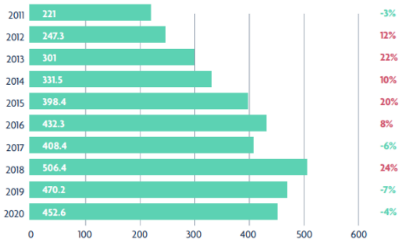

Fig. 2. Increase in Remote purchase fraud losses on UK issued cards [5]

Chip-based cards have reduced the prevalence of counterfeit card fraud, which involves utilizing a replica set of real card information that has been stolen as per fig. 1. According to FFA, credit card fraud appears to be occurring more frequently through online transactions, and 2016 statistics from the UK suggest a 20% increase over the prior year's findings, as shown in fig. 2. However, due to this COVID-19, it slightly decreased in 2019/2020. As shown in fig. 1, there are the following types of credit card fraud:

- Lost/Stolen credit card: A physical credit card is stolen from the wallet or mail of the cardholder and used to make fraudulent purchases.
- Card-not-present fraud: When making a purchase, a fraudster uses a credit card that has been stolen or compromised, but the card is not physically present (e.g., online or over the phone).
- Counterfeit fraud: This occurs when a culprit generates a false credit card with stolen or created information and uses it to make fraudulent purchases.
- Remote purchase credit card fraud: When a credit card is used for an online purchase or a phone order and the card is not physically present at the time of the transaction, it constitutes a sort of credit card fraud.
- Card ID Theft: This happens when someone's identity is stolen and used to make fraudulent purchases or apply for credit. The information stolen includes names, addresses, and credit card numbers.

Fraud prevention involves implementing various procedures, safety measures and controls to reduce the likelihood of fraudulent activities occurring. This proactive approach aims to thwart fraud before it happens, thereby minimizing potential losses and damages to both individuals and businesses. Detection, on the other hand, focuses on identifying fraudulent activities that may have already occurred, often through monitoring transactions and analyzing patterns for suspicious behaviour. Both prevention and detection are crucial components of effective fraud management strategies.

Machine learning is a sub-field of computer science that deals with artificial intelligence where it has abilities to improvise prediction without changing the model based on previous data [6]. Machine learning is employed in this study to identify credit card fraud. The fraudulent transaction or payment done by an unauthorized card holder is known as Credit card fraud [7]. According to the Federal Trade Commission (FTC), there were around 1579 data breaches totalling 179 million data pieces, with credit card theft being the most common [8]. As a result, implementing an effective credit card fraud detection mechanism that can safeguard customers from financial loss is critical. Credit card fraud detection has low detection accuracy and is unable to deal with the highly skewed nature of credit card fraud datasets. As a result, it is critical to develop optimal feature selection that can function efficiently and detect credit card fraud with a high accuracy score [9]. More than 99% of all transactions are typically legitimate, meaning that less than 1% of them are fraudulent [10], [11]. Imbalanced datasets often lead to biased models that favour the majority class [8]. Ensemble techniques, especially those that incorporate techniques like resampling (e.g., boosting or bagging), can mitigate this bias by giving more weight to minority class instances or adjusting decision boundaries accordingly. Ensembles typically yield higher performance metrics such as precision, recall, F1-score, and area under the ROC curve (AUC-ROC) compared to individual classifiers, especially when dealing with imbalanced datasets. This is crucial in fraud detection scenarios where correctly identifying fraudulent cases (precision) while minimizing false negatives is paramount. The missing fraudulent transaction (false negatives) can be much more costly than falsely accusing legitimate transactions as fraud (false positives) in the fraud detection system. However, the homogeneous ensembles typically use multiple instances of the same base classifier with variations in training data or parameters, which may not offer sufficient diversity in model predictions. Heterogeneous ensembles combine classifiers that utilize different algorithms, representations, or even feature sets. This diversity allows the ensemble to capture a wider range of patterns and relationships in the unseen data. In a heterogeneous ensemble, one classifier might be better at detecting global patterns (e.g., SVM), while another might excel in capturing local patterns or outliers (e.g., k-NN). By aggregating predictions from diverse models, heterogeneous ensembles can leverage these complementary strengths to improve overall prediction accuracy and robustness. In imbalanced datasets, homogeneous ensembles may amplify biases present in the base classifier. Heterogeneous ensembles, by incorporating classifiers with different biases and strengths, can mitigate these biases and provide more balanced predictions across different classes, leading to improved performance metrics such as recall, precision, and F_β-score.

This paper aims to contribute to fraud detection on credit cards by developing feature selection and a classification model to capture a wider range of patterns and relationships in the unseen data. The paper's main objective is to create heterogeneous ensemble classification models for credit card fraud detection by selecting features. Feature selection is crucial in credit card fraud detection to enhance model performance and reduce computational complexity. Using the k-best features approach involves selecting the k most informative features from the dataset. The chi-squared statistic, mutual information, and f-value are a few typical scoring metrics in k-best selection. The heterogeneous ensemble consists of k-Nearest Neighbors (k-NN), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) to capture diversity effectively. In the ensemble, k-NN can contribute by capturing local patterns and anomalies within the data. SVMs contribute diversity by focusing on finding global decision boundaries that maximize the margin between fraud and legitimate transactions, particularly useful when fraud patterns are not linearly separable. ANNs contribute diversity by their ability to automatically learn and adapt to the data, extracting abstract features that may not be apparent to other classifiers like k-NN and SVM. ANN can model complex fraud patterns to enhance the ensemble's ability to detect subtle and evolving fraud behaviours. By leveraging the diversity of k-NN, SVM, and ANN in a heterogeneous ensemble for credit card fraud detection enhances the ensemble's ability to capture a wide range of fraud patterns and improve detection performance, making it a robust approach for complex and dynamic fraud detection tasks.

The major contributions of this paper can be summarized as follows:

1. Select k best relevant features based on statistical tests to improve the recall rate of the minority class.
2. Ensemble methods utilize the diversity of heterogeneous classifiers to enhance and robust overall fraudulent detection.

The rest of the paper is organized as follows: Section 2 discusses the related works. Section 3 includes the details of the entire process and the model used. Section 4 discusses the detailed results. Finally, a conclusion is in section 5.

II. Related Work

A. *Feature Selection*

Pattern recognition algorithms in analytics and machine learning face difficulties due to the massive increase in the amount and complexity of data produced by diverse applications. This problem can be solved using dimensionality reduction approaches, namely by using feature selection (FS) and feature extraction (FE). FS methods reduce data burden and help prevent model overfitting, but must adapt to deal with the dynamic and fast-growing nature of contemporary data [12]. The t-statistic is used for feature subset selection, where 18 and 10 features were selected in two cases respectively. The results indicated that the Probabilistic Neural Network (PNN) performed the best, which is followed by Genetic Programming (GP) [13]. The performance of heart disease prediction is enhanced with fewer features using feature selection across all models. The highest accuracy is achieved using backward feature selection and a decision tree classifier [14]. A GA-based feature selection method is proposed and implemented with the RF applied to the European cardholders credit card transactions dataset and 5 optimal feature vectors are generated [15]. The GA is a search heuristic used to find approximate solutions but it is computationally expensive and time-consuming. Feature selection is recognized as an effective method to address imbalanced classification problems. This process can be formulated as a multiobjective optimization problem (MOP) that aims to identify a small subset of features while achieving high classification accuracy [16]. Two examples that use bootstrap samples from the training set are bagging and boosting [17], [18]. A bootstrap sample is an exact copy of the datasets produced by randomly choosing k occurrences and replacing them with new ones from the training set. Every replica is sent into a filter. Simple voting is used to combine the predictions of each classifier. The boosting strategy, on the other hand, samples the instances in accordance with their weights. Instances that the prior model incorrectly categorized are given more weight. For text classification issues, three often used ranker documents, such as frequency threshold, information gain, and chi-square were combined [19]. Additionally, feature selection techniques have been used in classification issues in the fields of signal processing and bio-informatics [20]. An ensemble based on GA-wrapped feature selection using three real-world data sets has outperformed a single classifier using all feature sets. The best individual classifier is then obtained from the GA-based feature subset selection [21].

B. *Classification Models*

The seriousness of credit card fraud has led to widespread recognition of and implementation of countermeasures. Banks and other financial institutions take pride in acting as the main barrier against fraud in addition to providing financial services to their clients. Furthermore, they invest in and develop various methodologies, including cutting-edge machine-learning techniques that many systems rely on extensively. Researchers have had to improvise with their research observations and patterns as cybercriminals have developed in their use of various distinct strategies. Countering fraud activities through data mining and machine learning is a well-known strategy to stop the damages brought on by these illegal acts. Data mining algorithms mainly examine the patterns and characteristics of legitimate and fraudulent transactions on data. On the other hand, machine learning techniques use classifiers to detect whether anticipated transactions are fraudulent or not. Pattern learning can be used to distinguish between authentic and fraudulent transactions by combining data mining and machine learning approaches[22].

The ensemble learning technique is proposed by combining random forest (RF) and ANN where RF provides higher accuracy and ANN detects fraud instances [2]. RF, Naive Bayes (NB), and Multilayer Perceptron machine learning-based techniques were used for detecting credit card fraud. The dataset covers transactions done by European credit card holders within the last two days. The researcher used the SMOTE oversampling technique to address the dataset's class imbalance problem [23]. A comparative study on credit card fraud detection is carried out but does not guarantee to give the same results in all environments. A high detection rate is offered by NN, NB, fuzzy systems and k-NN however, logistic regression, SVM and decision tree provide a low detection rate [24]. The cardholders are clustered into different groups based on the transaction amount, i.e., high, medium and low. Then using the sliding window strategy, aggregate the transactions made by the cardholders from aggregate the transactions into respective groups i.e. extract features to find cardholders behavioural patterns. These dynamic changes in parameters lead the system to adapt to new cardholders transaction behaviours timely [8]. In real card payment transaction, total of 13 statistical and machine learning models are applied to detect fraud using both publicly available and real transaction records. The results from both original features and aggregated features are analyzed and compared. A statistical hypothesis test is conducted to evaluate whether the aggregated features identified by a genetic algorithm can offer better discriminative power than the original features in fraud detection [25].

The fraud detection system is proposed using Kernel-based supervised hashing (KSH) by approximating the nearest neighbour search to identify and provide the most similar existing fraud samples for a transaction when it is predicted to be fraudulent. It is best suited for large high-dimension. The proposed model efficiently and accurately identifies fraudulent transactions and increases the overall efficacy of fraud detection systems [26]. A real dataset from a Turkish bank is used where the misclassification rate is reduced by Genetic Algorithm (GA) and scatter search [27]. Credit card fraud is detected using NB, SVM, and Deep learning on publicly available credit card datasets. Individual and hybrid models were assessed, employing AdaBoost and majority voting combination methods. The Matthews Correlation Coefficient (MCC) was used as the performance metric due to its consideration of true and false positive and negative outcomes [28].

III. Methodology

Fig. 3 shows the credit card fraud detection methodology using an ensemble technique based on k-NN, SVM and ANN.

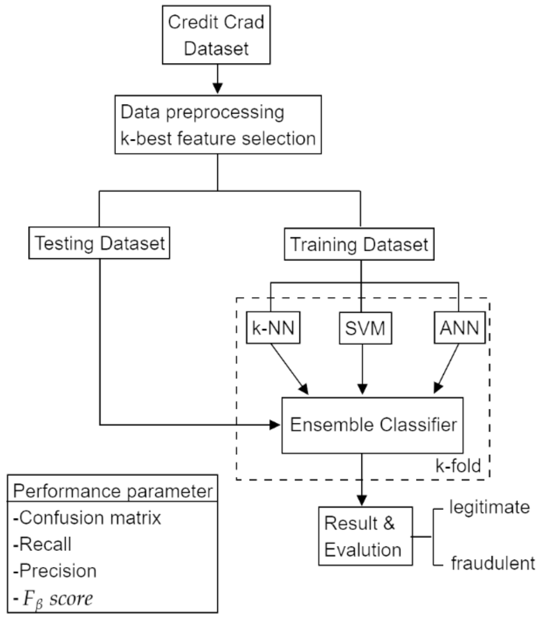

Fig. 3. Proposed block diagram of credit card fraud detection using heterogeneous ensemble techniques

In this research, the dataset for credit card transactions performed by European cardholders over two days in September 2013 is used [10]. From table 1 the dataset has a total of 284807 transactions, with 0.172% of them being fraudulent.

TABLE I. Credit Card Dataset

Normal	Fraudulent	Features	Instance
284,315	492	30	284,807

The key component obtained by principal component analysis (PCA) in this dataset has 30 features (V1,⋯, V28), as well as Time and Amount which the PCA is not changed. The class (i.e., fraudulent or legitimate) is represented by the last column, with a value of 1 denoting a fraudulent transaction and a value of 0 otherwise. For data security and integrity considerations, the characteristics V1 through V28 are unnamed[10]. This dataset was employed in [23] and one of the significant concerns observed was the low detection accuracy score produced by those models as a result of the dataset's extremely imbalanced nature. The definition variables used in the dataset are displayed in table 2.

The key to comprehending data is correlation matrices. The data can be represented graphically using the correlation between variables, where correlation denotes the interdependence of the variables. Typically, during the training phase, feature variables with greater correlations to the response variable have a more significant impact. A correlation matrix is depicted in fig. 4 and explains the pairwise correlation between each variable. The provided correlation matrix demonstrates that there is no association between any of the major components from V1 to V28. A closer look reveals that there is no association between "Time" or "Amount" and the response variable "Class." Nonetheless, there are some positive and negative correlations between the principal components and the "Class" variable. Moreover, a negative correlation exists between V17, V14, V12, and V10. Observe how the likelihood of a fraudulent transaction occurring increases as these values decrease where a Positive correlation exists between V2, V4, V11, and V19. Observe how the likelihood that a fraudulent transaction would occur increases as these values increase.

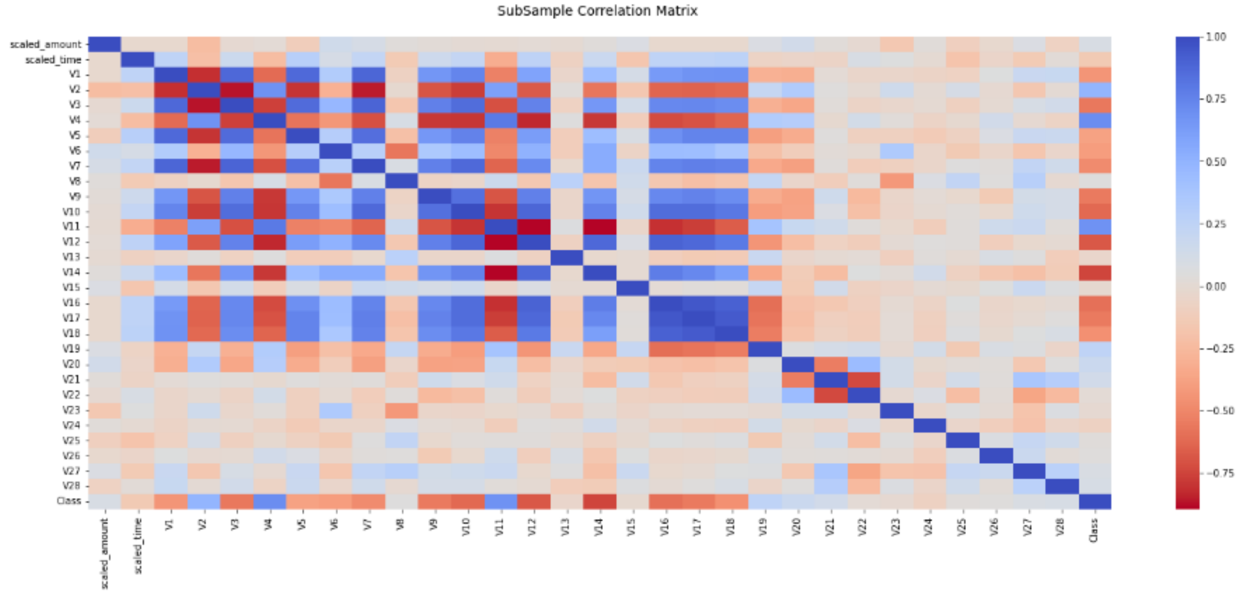

Fig. 4. Illustration of feature correlation of credit card dataset

A. Feature Selection

Feature selection and feature extraction are two standard methods to reduce the number of features used to characterize datasets and improve the performance of classifiers. FS selects a subset of the most relevant features from the dataset which should be informative and discriminating. A complementary area of research to FS is Feature Ranking (FR), which involves scoring each feature according to a particular criterion, then selecting the top k features concerning this score (k-best feature selection) [29].

TABLE II. DESCRIPTION OF THE VARIABLES IN THE DATASET

Variable	Type	Description
Time	Integer	Time elapsed between each transaction and the first transaction
Amount	Double	Transaction Amount
Class	Integer	Response variable (1=Fraudulent and 0=Genuine)
V1	Double	First Principal Component
V2	Double	Second Principal Component
V3	Double	Third Principal Component
…	…	…
V28	Double	Last Principal Component

Each feature is assigned a score based on a statistical test such as chi-squared, F-value (ANOVA F-test) or a measure of its relevance such as mutual information. The k- best method is used due to its computationally efficient but may not capture feature interactions since it evaluates each feature independently. The k-best method can be employed to select the most relevant features based on different scoring functions such as mutual information and f-test to classify problems.

The mutual information measures the information gained about the target variable **Y** from each feature **X**, capturing both linear and non-linear relationships. Mutual information quantifies the reduction in uncertainty about **Y** due to the knowledge of **X**. The mutual information can be written as:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ H(Y) &= -\sum_{y\in Y} P\,(y)\log P(y) \\ H(Y|X) &= -\sum_{x\in X}\sum_{y\in Y} P\,(x,y)\log P(y|x) \end{aligned} \qquad (1)$$

where $H(Y)$ is the entropy of the target variable **Y**. $H(Y|X)$ is the conditional entropy of **Y** given feature **X**. The F-test is used for classification problems where the target variable is binary (fraud or not fraud). It evaluates whether there is a significant difference between the means of the two classes for each feature. The F-statistic test shows whether there is a significant difference between the means of the two classes and can be expressed as:

$$\begin{aligned} F &= \frac{MSB}{MSW} \\ MSB &= \frac{n_1(\bar{X}_{i1}-\bar{X}_i)^2+n_2(\bar{X}_{i2}-\bar{X}_i)^2}{k-1} \\ MSW &= \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n-k} \end{aligned} \qquad (2)$$

where MSB is the mean square between $k = 2$ class (i.e., fraud or not), $\bar{X}_i$ is the overall mean of the feature $\mathbf{X}_i$, n_1 and n_2 are the sample sizes for the two classes (i.e., fraud or not) and $\bar{X}_{i1}$ and $\bar{X}_{i2}$ are the means of the feature for the two classes. MSW is the mean square within groups (i.e., within each class), s_{i1}^2 and s_{i2}^2 are the variances of the feature within the two classes and n is the total number of transactions.

B. Classification Base Model

This research aims to capture a wider range of patterns and improve detection performance in a changing behavioural fraud pattern environment. k-NN captures local patterns and anomalies within the data. It finds the k-nearest data points (neighbours) to a given transaction and classifies the transaction based on the majority class of its neighbours. It does not make any assumptions about the underlying data distribution, making it flexible in capturing subtle and local anomalies that might indicate fraud. In k-NN, the distance metric calculates the separation between a new transaction **X** and each transaction $\mathbf{X}_i$ in the training set. The method operates by locating the k data points that are closest to a given point and utilizing those data points to forecast the point's class label. The distance between two points can be determined by the Minkowski distance metric as:

$$d(X,X_i) = \left(\sum_{j=1}^{n}\left|X_j - X_{ij}\right|^p\right)^{1/p} \qquad (3)$$

where p is the power parameter. The decision boundary's degree of "smoothness" or "ruggedness" is determined by the value of P. When P is smaller, the boundary will be rockier, and when P is larger, the boundary will be smoother.

- When $P = 1$, Manhattan Distance, also known as $L1$ Norm, is calculated as the sum of the absolute differences between the coordinates of the points.
- When $P = 2$, Euclidean Distance, also known as $L2$ Norm, is the most common distance metric and is calculated as the differences between the coordinates of the points.
- When $P \to \infty$, Chebyshev Distance, also known as $L\infty$ Norm or Maximum Distance, is calculated as the maximum absolute difference between the coordinates of the points.

In k-NN, the weighting function is a crucial hyperparameter that determines how much influence each of the k nearest neighbours has on the final decision. Commonly used weighting functions include uniform weighting and distance weighting.

- Uniform Weight assigns equal importance to each of the k nearest neighbours in deciding the output class. This means that the 1st nearest neighbour has the same influence on the decision as the k^{th} nearest neighbour.

$$Weight(x_k) = 1 \qquad (4)$$

- Distance Weight assigns a weight to each of the k nearest neighbours based on their distance from the input sample. The further a neighbour is, the less weight it carries, giving more importance to closer neighbours. This method is often used because closer neighbours are likely to be more similar to the input sample. Mathematically, distance weighting can be represented as:

$$Weight(x_k) = \frac{1}{d(x_k,x)} \qquad (5)$$

where $d(x_k, x)$ is the distance between the input sample x and the k-th nearest neighbour x_k.

SVMs focus on finding global decision boundaries that maximize the margin between fraud and legitimate transactions. By using kernel functions (e.g., radial basis function, polynomial), SVMs can handle cases where fraud patterns are not linearly separable, thus capturing more complex relationships in the data. The SVM identifies either a transaction that falls in a fraud or not by:

$$y = c\{\mathbf{w}^{\mathbf{T}}K(\mathbf{x},\mathbf{z}) + b\} + \frac{1}{2}\|\mathbf{w}\|_2^2 \qquad (6)$$

where $\mathbf{w}$ is weight vector, b is biased, c is regularization parameter and $K(\mathbf{x},\mathbf{z})$ is kernel such as linear, sigmoid, polynomial, radial basis function (RBF). The optimal kernel is used to classify the transactions by non-linearly separating the data. The SVM equation with i^{th} training value is expressed as:

$$\begin{aligned} \pi^+ &: \mathbf{w}_i\mathbf{x}_i + b = +1 \\ \pi^- &: \mathbf{w}_i\mathbf{x}_i + b = -1 \end{aligned} \qquad (7)$$

ANNs can automatically learn and adapt to the data, extracting abstract features that may not be apparent to k-NN and SVM classifiers. It can model complex relationships and interactions in the data, which makes it effective at detecting subtle fraud patterns that k-NN and SVM models might miss. ANN trains on historical transaction data, allowing it to learn the intricate patterns associated with fraudulent and non-fraudulent transactions. The cross function of ANN is given by cross-entropy loss for binary classification as:

$$L \;\; = -\frac{1}{C}\sum_{c=1}^{C}[y_c\log(\hat{y}_c) + (1 - y_c)\log(1 - \hat{y}_c)] \quad (8)$$

where $C = 2$ is the number of classes, i.e., 1 for fraud, 0 for non-fraud. $\hat{y}_c$ is the predicted probability of fraud for transaction c and y_c is the actual label.

C. Ensemble Classifier

The combination of k-NN, SVM and ANN in a heterogeneous ensemble for credit card fraud detection enhances the overall predictive performance by combining the strengths of these individual classifiers. In the ensemble method, a voting classifier is applied to aggregate predictions from multiple individual classifiers (k-NN, SVM, ANN) using a majority vote, i.e., hard voting or by averaging predicted probabilities, i.e., soft voting.

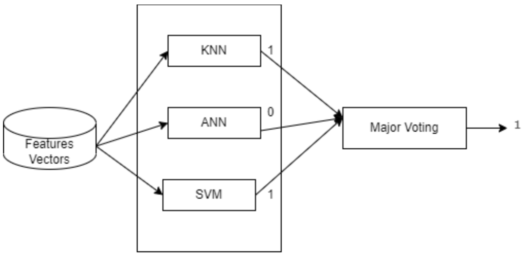

Fig. 5. Ensemble Hard Voting using k-NN, SVM and ANN

Fig. 5 shows the majority vote among base classifier, i.e., k-NN, SVM and ANN. The prediction for a given set of data is [1,1,0] from k-NN, SVM and ANN. Each classifier are assigned with equal weights, where 1 for fraud and 0 for not fraud. The prediction is made by mode of [1, 1, 0] is 1, and as a result, the predicted class for the specific record is changed to class 1.

$$mode[kNN, ANN, SVM]$$
$$mode[1,1,0] = 1$$

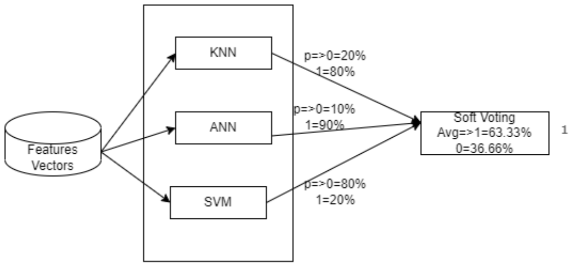

Fig. 6. Ensemble Soft Voting using k-NN, SVM and ANN

The final prediction in a soft voting classifier is based on the class that the base classifiers gave the highest probability as shown in fig. 6. It considers the probabilities that each classifier assigns to each class, rather than only looking at the majority vote of classifiers. For example, let's say there is a binary class [0,1] and the probabilities calculated by the classifier: k-NN [0.2,0.8], ANN [0.1,0.9], and SVM [0.8,0.2] With equal weights, i.e., $1/3$, the probabilities will be calculated as follows:

$$Class_0 = 0.33 \times 0.2 + 0.33 \times 0.1 + 0.33 \times 0.8 = 0.363$$

$$Class_1 = 0.33 \times 0.8 + 0.33 \times 0.9 + 0.33 \times 0.2 = 0.627$$

The ensemble classifier will estimate a probability of $[36.3\%, 62.7\%]$ as a result, the predicted class for the specific record is changed to class 1.

D. Performance Evaluation

Each class of credit card classification consists of an imbalanced dataset. It is important to identify fraudulent cases (precision) correctly while minimizing false negatives is paramount. The precision and F_β parameters are considered the main performance evaluation parameters of credit card fraud detection where β should be less than 1 (i.e., $\beta < 1$). The F_{β_c} score of the c-th class is calculated as:

$$F_{\beta_c} \;\; = \frac{(1+\beta^2)(\text{Precision}_c \times \text{Recall}_c)}{\beta^2 \times \text{Precision}_c + \text{Recall}_c} \qquad (9)$$

The plotted AUC for each model helps to categorize the quality of the model by showing how well a classifier performs a specific classification task.

IV. Results and Discussion

A. Without Feature Selection

The dataset is split into training and test sets with a ratio of 75:25. The ensemble classifier is used to the dataset without any feature selection and evaluated with precision and $F_{0.1}$ score where $\beta = 0.1$.

TABLE III. Classification Results without Feature Selection

Model	Accuracy	Precision	Recall	$F_{0.1}$
kNN	99.836	100	5.102	82.989
SVM	99.845	60.416	29.591	59.731
ANN	99.612	28.113	80.612	28.3171

Tab. 3 shows the classification report of the base model classifiers. k-NN and SVM classifiers have minimised the false positive, i.e., accusing legitimate transactions as fraudulent but unable to identify transactions as fraudulent. The ANN model has a high recall (80.6122%) but low precision (28.1139%). This indicates that while the model is good at identifying most fraudulent transactions but also identifies legitimate transactions as fraudulent. Tab. 4 shows the classification performance of a Voting classifier using hard and soft voting for credit card fraud detection without feature selection where the base model classifiers are k-NN, SNM and ANN. Both hard and soft voting show high precision for the fraud class (0.875 for hard voting and 1.000 for soft voting). This indicates that the model can predict a transaction as fraudulent, it is usually correct. However, the recall is significantly low for both voting types (0.072 for hard voting and 0.020 for soft voting). This indicates that the model is missing many actual fraudulent transactions.

TABLE IV. VOTING CLASSIFICATION WITHOUT FEATURE SELECTION

TABLE V. OTING TYPE	TABLE VI LASS	TABLE VII. RECISION	TABLE VI ECALL	TABLE I	TABLE X INAL $F_{0.1}$
Hard	***non-fraud***	0.998	0.999	0.998	0.780
	fraud	0.875	0.072	0.780	
Soft	***non-fraud***	0.998	1.000	0.998	0.658
	fraud	1.000	0.020	0.658	

The hard voting method has a higher final $F_{0.1}$ score (0.780) compared to soft voting (0.658). This suggests that hard voting is better at handling the fraud detection task in this scenario, given the higher weight placed on precision. It effectively combines the strengths of the individual classifiers to enhance precision. Hard and soft voting show excellent performance for the non-fraud class, with very high precision and recall values, indicating that the model correctly identifies non-fraudulent transactions almost all the time. Both methods excel in identifying non-fraudulent transactions but struggle with detecting all fraudulent transactions, highlighting the challenge of imbalanced datasets in fraud detection. The ensemble methods can enhance the precision of the model in comparison to the individual classifiers. However, the recall of the ensemble method does not improve credit card fraud detection. The performance of recall can be enhanced by selecting features that help identify as many fraudulent transactions as possible.

B. *With Feature Selection*

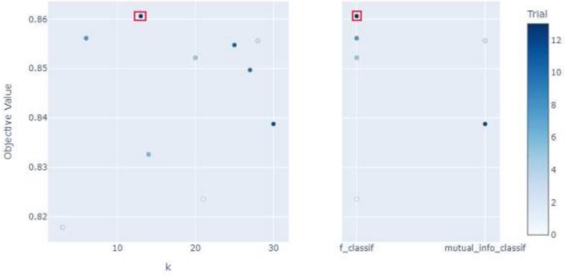

Fig. 7. Hyperparameter of k-best to select features of credit card fraud detection in terms of accuracy as an objective value.

Fig. 7 shows the performance of the k-best model with k optimal features and scoring functions (i.e., mutual information and F-test). The 13 optimal features (i.e., k=13) are selected out of 30 features of the data and the F-test provides the optimal performance than the mutual information in terms of accuracy as an objective value. Thus, the 13 optimal features using the F-test have been implemented to classify fraudulent credit cards which are [’V1’, ’V3’, ’V4’, ’V5’, ’V7’, ’V9’, ’V10’, ’V11’, ’V12’, ’V14’, ’V16’, ’V17’, ’V18’]. Then the 13 feature sets generated by the k-best are used to detect credit card fraud using classifiers k-NN, ANN, SVM and ensemble methods.

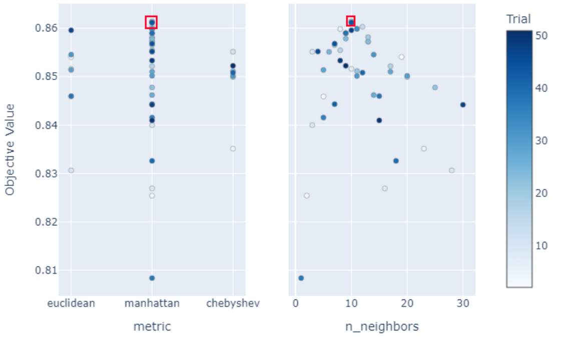

Fig. 8. k-NN hyperparameter tuning of k nearest neighbour with distance metric in terms of accuracy as an objective value

Fig. 8 and 9 show the hyperparameter tuning of classifier algorithm k-NN where the classifier identifies 10 nearest neighbours to the input data point based on their Manhattan distances. Distance weighting is assigned to each of the 10 nearest neighbours based on their distance from the input sample. The optimal accuracy is obtained at 8-fold cross-validation of k-NN.

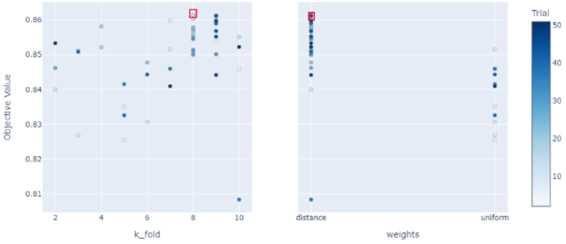

Fig. 9. k-NN hyperparameter tuning of weight function and k-fold cross validation in terms of accuracy as an objective value.

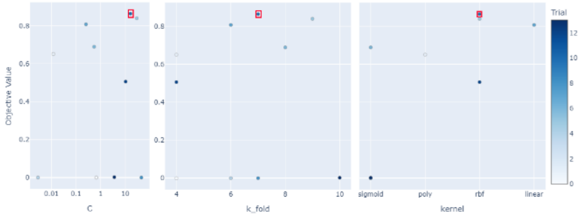

Fig. 10. SVM hyperparameter tuning of C, kernel and k-fold cross validation in terms of accuracy as an objective value.

The support vector machine is used to find global decision boundaries between fraud and legitimate transactions. Fig 10 shows hyperparameter tuning of the SVM where fraudulent and legitimate credit cards are separated in higher dimensions using radial basis function kernel and a tuned regularization parameter $C = 16.148$ to avoid over-fitting. The 7-fold cross-validation ensures the robustness of SVM model performance.

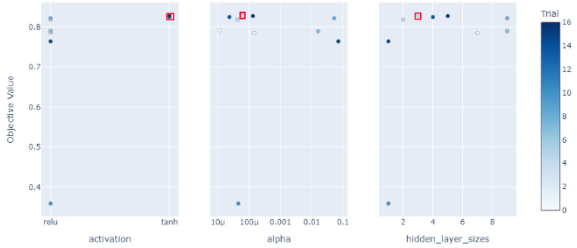

Fig. 11. ANN hyperparameter tuning of activation function, learning rate (\alpha) and number of hidden layers in terms of accuracy as an objective value.

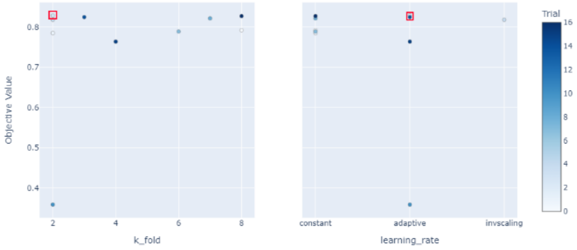

Fig. 12. ANN hyperparameter tuning of k-fold cross validation and types of learning rate in terms of accuracy as an objective value.

ANN is used to automatically learn and adapt to the data, extracting abstract features. Fig 11 and 12 show the hyperparameter tuning of the ANN classifier with 3 hidden layers and a tanh activation layer. By using tanh activation function, the model effectively captures complex patterns in the data, which is essential for distinguishing between fraudulent and legitimate transactions. The optimal learning rate 6.216×10^{-05} is initiated to control how much the model's weights are updated with each step. An adaptive learning rate adjusts during training to improve convergence without overshooting the minimum loss. Three hidden layers provide a balance between learning complex patterns and avoiding overfitting which improves its predictive power.

TABLE XI. CLASSIFICATION RESULTS WITH K SELECTED FEATURE

Model	Accuracy	Precision	Recall	$F_{0.1}$
kNN	99.948	92.195	76.829	91.992
SVM	99.937	86.511	75.609	86.374
ANN	99.936	86.4486	75.203	86.306

Tab. 5 shows the classification results of credit card fraud detection models using k-best optimal feature selection with hypertuned base classifiers, i.e., k-NN, SVM and ANN. The optimized hyperparameters for each model to enhance performance with 13 selected features using k-best feature selection. Recall measures the model's ability to correctly identify fraudulent transactions (true positives). KNN has the highest recall, indicating that it is slightly better at detecting fraud compared to ANN and SVM. Higher recall is crucial in fraud detection to minimize the number of fraudulent transactions that go undetected. Precision measures the accuracy of the fraud predictions (how many identified frauds are actually frauds). KNN also leads in precision, suggesting that it not only captures more frauds but does so with fewer false positives compared to ANN and SVM. Here, KNN has the highest $F_{0.1}$ score, indicating that it provides a better balance between precision and recall compared to the other models.

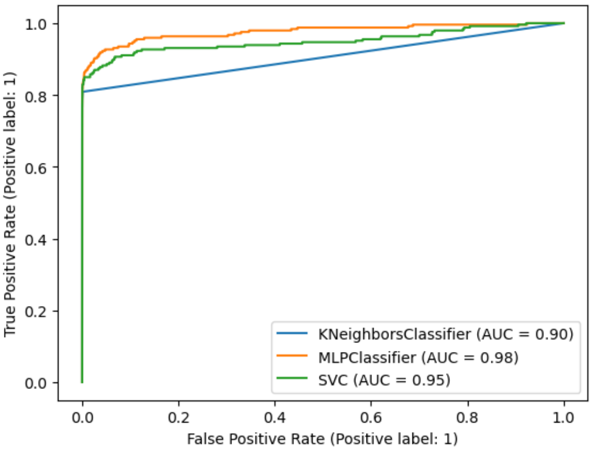

Fig. 13. AUC results with k-optimal selected features

Fig. 13 shows the ROC curve plot compares the performance of three classifiers (k-NN, SVM and ANN) in detecting credit card fraud. The AUC (Area Under the Curve) values indicate the overall performance of each model. ANN has the best performance with an accuracy of 0.98, making it the most suitable model among the three for detecting credit card fraud. The performance of the baseline model classifier can be improved by combining these models with an ensemble method like a voting classifier to leverage their strengths.

The tab. 6 presents the classification report for Hard and Soft voting methods used in the ensemble model for detecting credit card fraud. Hard Voting shows better recall for the fraud class, making it more effective in identifying actual fraud cases. This is crucial in fraud detection, where missing fraudulent transactions can have severe consequences. Soft Voting exhibits slightly better precision and overall $F_{0.1}$ score, suggesting it is slightly more reliable in minimizing false positives.

TABLE XII. CLASSIFICATION REPORT OF HARD VOTING AND SOFT VOTING

Voting	Class	Precision	Recall	$F_{0.1}$
Hard	***non-fraud***	0.999	0.999	0.999
	fraud	0.930	0.806	0.928
Soft	***non-fraud***	0.9995	0.9998	0.999

Voting	Class	Precision	Recall	$F_{0.1}$
	fraud	0.934	0.775	0.932

V. Conclusion and Future works

Credit card fraud is a significant issue in financial sectors, necessitating advanced detection methods to minimize financial losses and improve security. Traditional methods often struggle with the high dimensional of transaction data and the imbalanced nature of fraud datasets. This paper proposes an ensemble learning method that leverages the diversity of k-NN, SVM, and ANN classifiers, combined with k-best feature and voting methods, to effectively identify fraudulent transactions. k-best feature selection method has also been utilized to improve classifier performance by reducing dimensional and focusing on the most informative features. The selected features are used to classify fraudulent transactions using base classifiers. The outcome of classifiers is enhanced by combining the base classifiers with soft voting. Therefore, the proposed model can obtain a higher minority class recall rate while also maintaining the minority class precision rate. The ensemble performs better than existing models in terms of accuracy, precision, recall, and $F_{0.1}$ score metrics. Thus, the proposed method significantly improves classification performance by leveraging diversity, making it a substantial performance improvement in detecting fraudulent credit cards. The ensemble method effectively balances the trade-offs between precision and recall, particularly in improving the recall rate, which is critical in identifying fraudulent transactions. The promising results from this study suggest that integrating advanced feature selection and ensemble learning can substantially improve fraud detection systems, providing a robust framework for financial institutions to safeguard against fraudulent activities.

References

[1] “Imposter Scams Top Complaints Made to FTC in 2018.” 2019.

[2] I. Sohony, R. Pratap, and U. Nambiar, “Ensemble learning for credit card fraud detection.” pp. 289–294, 2018.

[3] G. Babatunde Iwasokun, “Encryption and Tokenization-Based System for Credit Card Information Security,” *International Journal of Cyber-Security and Digital Forensics*, vol. 7, no. 3, pp. 283–293, 2018.

[4] Europol, *Situation report - payment card fraud 2012*.

[5] U. Finance, “FRAUD - THE FACTS 2021 THE DEFINITIVE OVERVIEW OF PAYMENT INDUSTRY FRAUD.” 2021.

[6] A. Burkov, *The hundred-page machine learning book*. 2019, pp. 3–5.

[7] S. Maniraj, A. Saini, S. Ahmed, and S. Sarkar, “Credit card fraud detection using machine learning and data science,” *International Journal of Engineering Research*, vol. 8, no. 9, pp. 110–115, 2019.

[8] V. N. Dornadula and S. Geetha, “Credit card fraud detection using machine learning algorithms,” *Procedia computer science*, vol. 165, pp. 631–641, 2019.

[9] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, “Real-time credit card fraud detection using machine learning,” in *2019 9th international conference on cloud computing, data science engineering (confluence)*, 2019, pp. 488–493.

[10] “Credit card fraud detection.” *https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud, 2013*.

[11] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, “Credit card fraud detection: A realistic modeling and a novel learning strategy,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784–3797, 2018.

[12] B. Venkatesh and J. Anuradha, “A review of feature selection and its methods,” *Cybernetics and information technologies*, vol. 19, no. 1, pp. 3–26, 2019.

[13] “Detection of financial statement fraud and feature selection using data mining techniques,” *Decision Support Systems*, vol. 50, no. 2, pp. 491–500, 2011.

[14] K. Dissanayake and M. G. Md Johar, “Comparative study on heart disease prediction using feature selection techniques on classification algorithms,” *Applied Computational Intelligence and Soft Computing*, vol. 2021, no. 1, p. 5581806, 2021.

[15] E. Ileberi, Y. Sun, and Z. Wang, “A machine learning based credit card fraud detection using the GA algorithm for feature selection,” *Journal of Big Data*, vol. 9, no. 1, p. 24, 2022.

[16] S. Han, K. Zhu, M. Zhou, and X. Cai, “Competition-driven multimodal multiobjective optimization and its application to feature selection for credit card fraud detection,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 12, pp. 7845–7857, 2022.

[17] L. Breiman, “Bagging predictors.” Aug. 1996.

[18] R. E. Schapire, “A brief introduction to boosting,” in *Proceedings of the 16th international joint conference on artificial intelligence - volume 2*, in IJCAI’99. Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999, pp. 1401–1406.

[19] J. O. S. Olsson and D. W. Oard, “Combining feature selectors for text classification,” in *Proceedings of the 15th ACM international conference on information and knowledge management*, in CIKM ’06. Arlington, Virginia, USA: Association for Computing Machinery, 2006, pp. 798–799.

[20] L. Zhang, “Optimal feature selection using distance-based discrete firefly algorithm with mutual information criterion.” Feb. 2016.

[21] E. Yu and S. Cho, “Ensemble based on GA wrapper feature selection,” *Computers and Industrial Engineering - COMPUT IND ENG*, vol. 51, pp. 111–116, Sep. 2006.

[22] O. S. Yee, S. Sagadevan, and N. H. A. H. Malim, “Credit card fraud detection using machine learning as data mining technique,” *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 10, no. 1–4, pp. 23–27, 2018.

[23] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, “Credit card fraud detection - machine learning methods,” in *2019 18th international symposium INFOTEH-JAHORINA (INFOTEH)*, 2019, pp. 1–5.

[24] Y. Jain, N. Tiwari, S. Dubey, and S. Jain, “A comparative analysis of various credit card fraud detection techniques,” *International Journal of Recent Technology and Engineering*, vol. 7, no. 5, pp. 402–407, 2019.

[25] M. Seera, C. P. Lim, A. Kumar, L. Dhamotharan, and K. H. Tan, “An intelligent payment card fraud detection system,” *Annals of operations research*, pp. 1–23, 2021.

[26] Z. Li, G. Liu, S. Wang, S. Xuan, and C. Jiang, “Credit card fraud detection via kernel-based supervised hashing,” in *2018 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2018, pp. 1249–1254.

[27] “Detecting credit card fraud by genetic algorithm and scatter search,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 13057–13063, 2011.

[28] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar, and C. V. N. M. Praneeth, “Credit card fraud detection using machine learning,” in *2021 5th international conference on intelligent computing and control systems (ICICCS)*, 2021, pp. 967–972.

[29] D. V. Akman *et al.*, “K-best feature selection and ranking via stochastic approximation,” *Expert Systems with Applications*, vol. 213, 2023.