Received: 23 October 2023 | Revised: 14 February 2024 | Accepted: 6 May 2024

DOI: 10.1002/asmb.2868

RESEARCH ARTICLE

WILEY

# K-Fuse: Credit card fraud detection based on a classification method with a priori class partitioning and a novel feature selection strategy

**Mohammed Sabri**[1,2] | **Rosanna Verde**[1] | **Antonio Balzanella**[1]

[1]Department of Mathematics and Physics, University of Campania L. Vanvitelli, Caserta, Italy

[2]Department of Computer Science, Faculty of Sciences Dhar EL Mehraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

**Correspondence**
Rosanna Verde, Department of Mathematics and Physics, University of Campania L. Vanvitelli, Caserta, 81100, Italy.
Email: rosanna.verde@unicampania.it

## Abstract

Online transactions have become the dominant and most popular form of online payment in today's digital economy. Due to the growing popularity of e-commerce and the convenience it offers, both consumers and businesses are rapidly adopting online transactions. Notably, credit cards have become one of the most popular and standard online payment methods. However, it should be noted that credit card transactions are not without challenges. In particular, detecting and preventing fraudulent transactions is a major concern of the online payment system. It is difficult to find an effective detection model that can detect the new patterns created by fraudsters, due to the constant evolution of their methods to exploit the vulnerability of current security protocols. These fraud patterns are evolving and may not correspond to existing documented models, leading to a reduction in their identification. In addition, the customer's behavior can affect the model detection as it is susceptible to change based on factors such as economic conditions, trends, and individual circumstances. When consumers deviate from their typical behavior, the model may generate false alerts, thereby reducing its ability to differentiate between legitimate and fraudulent transactions. This article presents a new supervised detection model, called K-Fuse, which introduces an unsupervised phase in order to detect fraud patterns that may correspond to innovative models introduced by fraudsters. K-Fuse is a supervised classification method that fuses three steps consisting of *(i)* unsupervised clustering to identify hidden patterns of transactions in a dataset, *(ii)* a novel feature selection criterion based on the unsupervised results, and *(iii)* supervised classification to exploit the results of clustering and feature selection to predict new transactions as fraudulent or legitimate.

KEYWORDS

credit card fraud detection, feature selection, supervised learning, unsupervised learning

All authors contributed equally to this study.

[Correction added on 26 September 2025, after first online publication: The copyright  line was changed.]

## 1 | INTRODUCTION

In recent years, e-commerce and online shopping have attracted large numbers of people around the world. The continued expansion of e-commerce on the Internet has raised new concerns, particularly regarding the increasing prevalence of fraud activities. Due to credit card fraud, financial institutions lose billions of dollars each year. The effect of these frauds affects both the victims and the credit card transaction companies.[1] Machine learning (ML) provides multiple solutions in terms of supervised or unsupervised algorithms for solving fraud issues. It is considered an efficient field for dealing with fraudulent transactions, in contrast to the classic fraud detection system that was a time-consuming and costly process because it analyzed only the most dangerous transactions alerted by a credit card company's investigation, and verified only a few alerts each day and the rest of the transactions classified as non-fraudulent until a customer flagged them as such.[2]

Most fraud detection and prevention systems (FDPS) in the banking sector are designed to monitor financial transactions and account activity in real-time to detect and prevent fraudulent activities such as unauthorized transactions and other financial crimes. FDPS is designed with a comprehensive strategy that includes both automated and human-controlled controls. The system employs ML models to automatically analyze large amounts of data in real-time, thereby identifying fraudulent activities in a timely manner. To further improve accuracy and prevent false positives, FDPS also includes human supervision to ensure that legitimate transactions are not rejected incorrectly. Although supervised methods are simple and their results more interpretable, they face two significant obstacles when used exclusively for fraud detection. First, they are unable to detect new types of fraud that fraudsters manipulate to evade detection, and second, when customers exhibit behavior that deviates from their typical patterns, they can result in false positives, which include legitimate transactions misidentified as fraudulent, or false negatives, which occur when fraudulent transactions are not identified appropriately; therefore, fraud detection systems can use unsupervised methods to extract new patterns in addition to supervised methods to address these challenges.

The main objective of this study is to enhance the identification of fraudulent activities by developing a model that not only recognizes existing fraud patterns but also uncovers novel vulnerabilities before they manifest as recognized fraud classes. Given the constantly changing and complex nature of today's financial crime, it is crucial to adopt a proactive approach. Fraudulent companies consistently modify their tactics to avoid being detected. Our approach utilizes historical data, which includes hidden patterns of fraudulent activity inside the dataset utilized for training. We suggest that these patterns can significantly influence the model's ability to detect new instances of fraudulent transactions. Our methodology aims to enhance the predictive precision of our model in identifying emerging fraudulent activities by doing research and deriving valuable insights from past instances of fraud. Our approach is specifically built to identify new patterns in consumer behavior that may indicate both genuine transactions and novel fraudulent tactics. Identifying such variances helps create new subgroups in the data, which improves our ability to distinguish between legitimate and fraudulent transactions more accurately. Our approach directly tackles the significant shortage in fraud detection, effectively tackling the constantly evolving issues and ensuring that our techniques stay efficient in preventing both existing and developing threats.

To improve credit card fraud detection, this article proposes a new three-step detection model. In the first step, a clustering method is used to generate subgroups from the a priori classes of the training set, using a new objective function that takes into account the compactness within subgroups and the separation between subgroups. This new criterion allows for an improvement in the performance of the classification algorithm in detecting fraud. The criterion is based on a weighted Euclidean distance to calculate the distance between transaction instances and the representative of their subgroup. In the second step, a new method of feature selection is introduced, which uses the feature weights discovered in the clustering step to measure the importance of the features.

In the third step, transactions are ranked using a detection model that relies on subgroups and their representatives derived from the clustering results and predictive features from the selection step.

Unsupervised classification detection techniques do not require knowledge of the transaction label and aim to group data into homogeneous subgroups.

The first step of the model is based on the assumption that hidden subgroups exist within the a priori classes, which need to be discovered before proceeding to the supervised classification step. It is worth noting that this step is a preprocessing step of the classification algorithm. Clustering based on new objective functions allows the identification of new patterns for which a specific predictive model should be used to label the data in the original groups.

By integrating the weights system into the objective function, it is possible to evaluate the importance of predictive features during the classification procedure, mainly when dealing with high-dimensional data. By employing a novel criterion to leverage this weight system, we can effectively identify the most pertinent features.[3] In addition, weights are assigned based on subgroup characteristics and optimization results of the objective function.

Supervised classification requires a dataset of past transactions whose labels are known. Based on this historical data, the main aim is to train a classifier algorithm that can determine whether a new transaction is fraudulent or not. In addition, for the sake of consistency, we propose a new strategy that assumes as input of the classifier, the results of the clustering and feature selection steps for representing subgroup labels and the most representative features. In addition, a new method will be implemented to ensure that the classification matches the labels of the original class.

In addition, in fraud detection, one problem that arises is that of class imbalance between fraudulent and non-fraudulent datasets. In our proposal, this aspect is considered and the balancing is done on the training set while the performance evaluation of the algorithm is considered on the imbalanced test set, precisely to take into account the characteristics of the data, that is the limited number of frauds compared to legitimate transactions for prediction purposes.

The main contributions and objectives of the work include the following results through the three steps:

- Step 1: Unsupervised clustering, to discover the hidden patterns in the original groups;
- Step 2: Feature selection, to reduce the features used before training a ML model to improve its performance in terms of execution time and efficiency;
- Step 3: Supervised classification, to detect the label of new transaction based on the results of the two previous steps.

The rest of the article is structured as follows: the Section 2 introduces related works, focusing on fraud detection methods encountered in the literature. The Section 3 illustrates the proposed approach to credit card fraud. Especially in Section 3.2 are specified the used resampling techniques to balance the training set. The Section 4 presents results obtained on two data sets. The Section 5 concludes the article and provides appropriate directions for future research.

## 2 | LITERATURE REVIEW

Numerous credit card transaction datasets containing legitimate transactions and fraud are available, both real and simulated. Apart from those subject to privacy regulations by financial institutions, many others are available in open repositories (e.g., data.world, Kaggle). These datasets are a valuable resource for the research community studying credit card fraud detection (CCFD). They contain an abundance of information about credit card transactions that can be used to develop ML models for fraud detection. In particular, supervised classification techniques have proven to be very effective in fraud detection, where historical transactions are used to train a classification model to determine whether a transaction is fraudulent or not.

Some of the notable methods identified to help detect fraud in credit cards include random forest (RF),[4,5] artificial neural network (ANN),[6,7] logistic regression,[8,9] support vector machine (SVM),[10,11] K-nearest neighbors (KNN),[12,13] and other techniques that have a hybrid and privacy-preserving approach for data privacy.

In Asha and Suresh Kumar[14] the authors used ML algorithms such as an ANN, KNN, and a SVM to predict fraud. Then they applied deep learning and supervised ML techniques to classify legitimate and fraudulent transactions.

In Xuan et al.[15] it is proposed the fraud detection based on the RF classifier of a Chinese e-commerce dataset. In this study, an improved RF classifier is also presented and the problem of class imbalance between fraud and non-fraud datasets is discussed.

Unsupervised classification techniques can be used to detect patterns or anomalies in credit card transactions that could indicate fraudulent behavior. These potentially allow the discovery of new fraud patterns that were previously undetected.[16,17]

Some problems encountered when dealing with credit card datasets include the high dimensionality and imbalance of the classes,[18,19] which make learning ML classifiers and accurate predictions difficult. Highly unbalanced credit card fraud datasets cause many traditional classifiers to fail to detect minority-class objects.[20,21] In addition, high-dimensional data often make the learning process complex and computationally expensive, yielding models with poor generalization capability.[22,23]

Therefore, feature selection is essential in such datasets to reduce the computational cost and enhance the model's generalization capability. The most important feature selection techniques are the filter and wrapper methods.

Filter methods rank the features on the basis of their individual contribution, regardless of the learning algorithm used for prediction, and can select a subset of features, or feature by feature based on a ranking method, without considering the classifier algorithm. While the wrapper feature selection methods have been widely applied in numerous applications.[24,25]

Wrapper methods use a ML algorithm to evaluate the performance of different subsets of features. These methods use a criterion to select the subset of features with the highest classification performance. Forward selection and backward selection are the most common wrapper methods. Forward selection starts with only a feature and proceeds by adding a feature at a time until the addition of an ulterior feature does not worsen classification performance. In contrast, forward selection, and backward selection, start with all features, or a set of features, and remove them one at a time until feature omission worsens classification performance. For example, Rtayli and Enneya[18] has experienced an increase in the classification performance of ML classifiers after the introduction of feature selection. In general, feature selection methods are useful in applications where the number of features affects the performance of the classifier. Ravisankar et al.[26] proposes a simple statistical technique that uses the t-statistic to perform feature selection on the dataset, identifying the most meaningful financial items that could detect the presence of fraud in balance statements.

The integration of supervised and unsupervised techniques has already been discussed in the fraud detection literature, and most research is limited to the use of unsupervised techniques to detect outliers in the data before using supervised classification.

In Wang et al.[27] the authors first partitioned and clustered the data set. Then, a RF framework, with C4.5 as the base classifier, was trained using the resulting clusters. The final result was determined by the majority vote.

In Carcillo et al.[28] propose a hybrid approach that combines unsupervised outlier scores and supervised learning for credit card fraud detection in order to solve the problem of fraudsters' ability to invent new fraud patterns, the benefit of using unsupervised outlier scores is to compute the outlier scores for different levels of granularity based on clustering analysis, their results on real dataset reported an improvement in terms of accuracy and enhances the detection performance of the classier model.

In Carcillo et al.,[28] a hybrid approach combining unsupervised outlier scores and supervised learning for credit card fraud detection is proposed to solve the problem of fraudsters' ability to design new fraud patterns. The advantage of using this approach is to calculate outlier scores for different levels of granularity based on clustering analysis.

In Eshghi and Kargari,[29] on the other hand, the authors have argued that unsupervised methods such as clustering and outlier detection techniques may not be adequate for complex fraud detection tasks and have proposed a framework with multi-criteria decision analysis and intuitionistic fuzzy sets to incorporate the effect of behavioral uncertainties to model the susceptibility of a banking transaction to be a fraud.

## 3 | THE K-FUSE MODEL

This section introduces the basic elements of the proposed model, K-Fuse, by integrating three distinct techniques: a new variant of the K-means algorithm, which introduces a partitioning step in the initial stage to group the classes of the training set into homogeneous subgroups; a revised version of the KNN classification algorithm, which uses clustering results to improve the performance of the classification algorithm; and a new feature selection criterion and sampling technique, to obtain a balanced training dataset. Additionally, we utilize a K-Fuse based approach to improve zero-shot learning, enabling the accurate identification of previously unseen fraud patterns in evolving datasets.

### 3.1 | Data exploration and visualization

In this subsection, we illustrate two real-world data sets that are being used to corroborate the K-Fuse model.

The first dataset (CD1) has been generously provided by our European financial partner, a major transnational services company. It contains 96 attributes, including the time at which a transaction occurred, the number of transactions, the type of transactions, the country in which the transactions take place, and other attributes. This dataset contains a total of 5 million transactions executed from March 1 to June 30, 2016, over a period of three months. In the experiments on such dataset, in order to preserve the confidentiality of client data and ensure the relevance of the predictions, we have limited the analysis to a selection of 75 variables, labeled F1 to F75, for legal reasons. The rows represent the total number

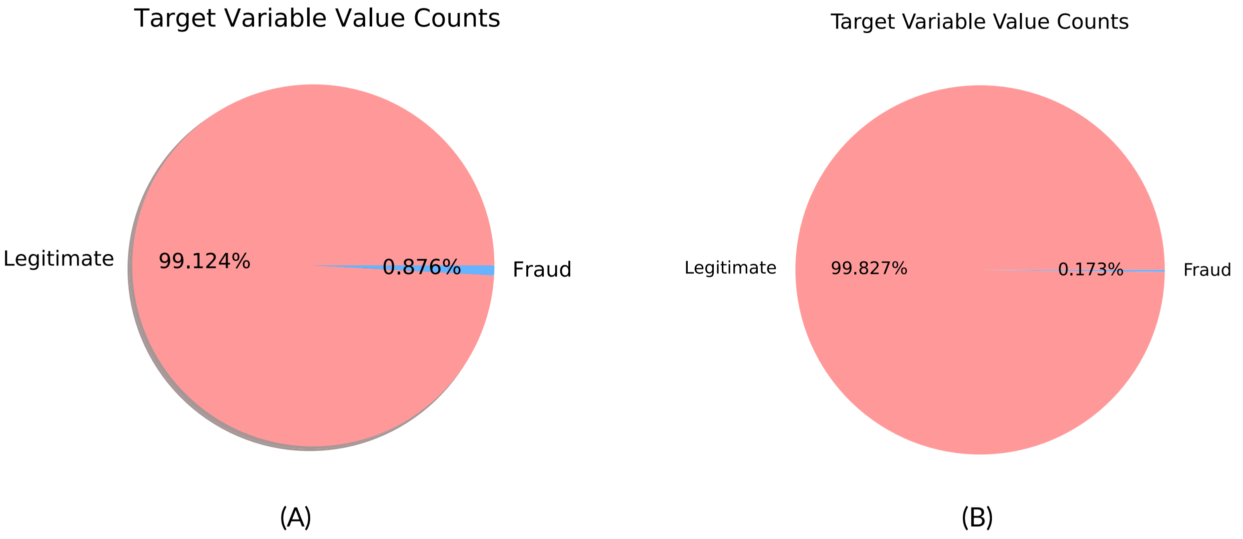

FIGURE 1 Target class ratio pie chart. (A) Target class ratio pie chart of CD1 dataset. (B) Target class ratio pie chart of CD2 dataset.

of transactions that have occurred. The percentage of transactions that correspond to illegal and fraudulent transactions amounts to 0.87 percent of the total dataset. To mitigate the issues associated with the test latency period, we chose to divide the data into a training set and a test set using a ratio, rather than relying on a specific time period of the data set.

The second dataset (CD2)* refers to credit card transactions made in September 2013 by European cardholders. It includes transactions that occurred over two days. Most of the variables in the input are numeric ones resulting from a transformation in PCA. For confidentiality reasons, the original features and additional context of the data cannot be revealed. The obtained principal components are represented as features $V1, V2, \dots, V28$, while the features “Time” and “Amount” remain the original ones. The “Time” feature indicates the elapsed time in seconds between each transaction and the first transaction in the dataset. The “Amount” feature, on the other hand, corresponds to the transaction amount and can be used for cost-sensitive learning. Collaborative research between Worldline and the Machine Learning Group at ULB (Université libre de Bruxelles), focusing on big data mining and fraud detection, collected and analyzed this dataset.

The target variable for both datasets considered is binary; it takes the value 0 for a legitimate transaction (negative label) and 1 for a fraudulent transaction (positive label).

In Figure 1, the pie chart shows the percentage distribution of the classes of the legitimate and fraudulent transactions in the two datasets. In Figure 2, the bar charts represent the frequency distributions of the classes in the two datasets. It can be seen in Figures 1 and 2, that the two classes are strongly imbalanced in both datasets; the number of frauds (anomalous transactions) is significantly lower than the number of legitimate transactions (normal transactions). Comparing the first and second datasets, the fraud rate is 0.87% and 0.17%, respectively. This large disparity between the classes (legitimate and fraudulent) may confuse our credit card fraud detection model, causing misclassification.

## 3.2 | Class imbalance in data classification

The problem of imbalanced datasets is very important in real-world applications such as medical diagnoses, detecting software deficiencies, financial issues, finding drugs, and bioinformatics.[30–33] The imbalanced data classification is still a major problem in the process of fraud detection. Indeed, with an imbalanced dataset, we find far fewer training instances of one class variable than the other class variable. Correspondingly, the first is known as the minority set and the second is known as the majority set.[34] When running an imbalanced dataset to detect fraudulent transactions, most classification models classify the majority class with a very high accuracy rate and the minority class with a much lower one. This means that the proposed classification model has a poor performance in diagnosing the minority class samples. Most learning algorithms do not fit a strong imbalance in class representations in the dataset. Therefore, learning from an imbalanced dataset is a challenging research issue that needs to be considered.[35,36] Overall, most of the classifiers suffer from class imbalance, some more than others. The strategies used to tackle this issue operate at two different levels: some aim to solve the imbalance of the dataset beforehand whereas others want to adapt the learning algorithm to make it robust to

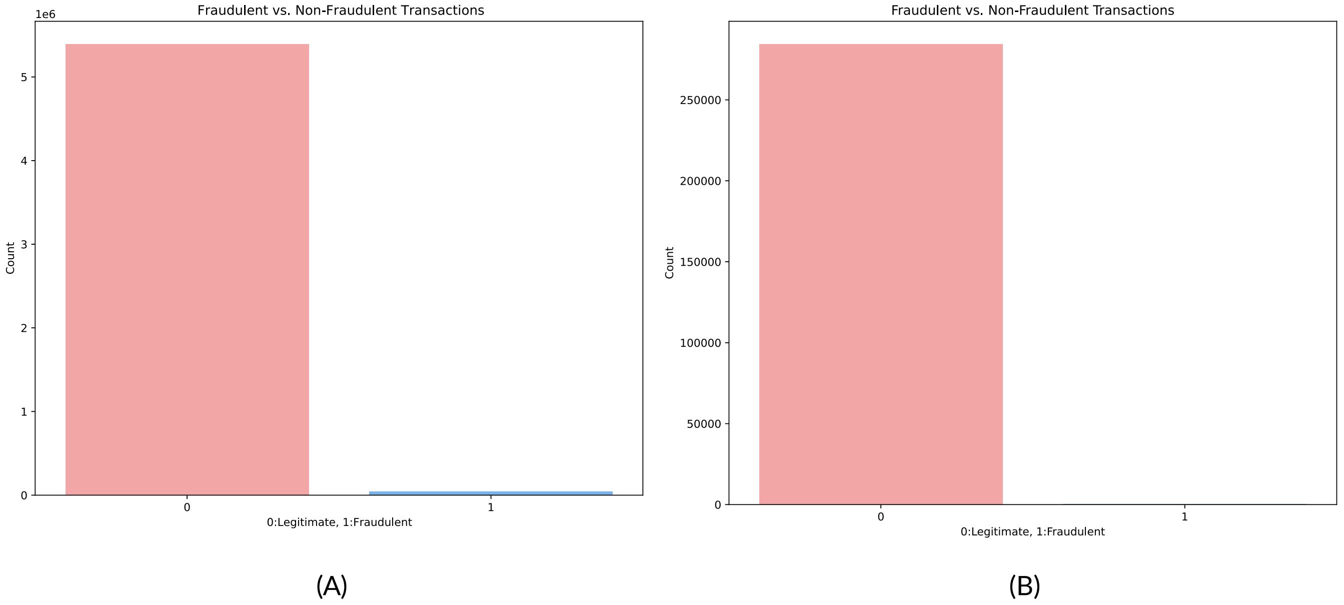

FIGURE 2 Histogram of counts for both classes across both datasets. (A) CD1 dataset. (B) CD2 dataset.

an imbalanced classification task.[37] The methods dealing with the problem of imbalanced datasets can be grouped into three categories[38,39]: majority class undersampling, minority class oversampling, or a mix of these ones.

The random undersampling technique involves reducing the size of the majority class by randomly removing observations until the dataset reaches equilibrium. In the context of an imbalance problem, it is often reasonable to assume that numerous observations belonging to the majority class are redundant. By randomly eliminating a portion of these observations, it is expected that the overall distribution of the data will not change significantly. Furthermore, the random undersampling technique is able to increment the learning process of a fraud detection algorithm.[40–42] According to Singh et al.,[21] the number of instances in the balanced dataset using RUS technique's algorithm can be written as follows:

$$|S| = S_{\min} + (S_{\max} - |R|), \tag{1}$$

where $S$ is a dataset with a class imbalance problem. In the dataset, let $S_{\min}$ be the number of examples of the minority class and $S_{\max}$ the number of instances of the majority class; $R$ is the randomly selected subset of the original dataset used to reduce the majority class and balance the two classes.

The synthetic minority oversampling strategy (SMOTE) is an oversampling method that generates synthetic samples from the minority class by leveraging the information contained within the existing data.[43] For each instance derived from the minority class. In contrast to the RUS technique, the calculation of the number of SMOTE samples in the balanced dataset can be expressed as follows:

$$|S| = S_{\max} + (S_{\min} + |R|). \tag{2}$$

Specifically, SMOTE algorithm is designed to address the issue by randomly selecting a data point from the minority class to serve as a reference point. In the context of the given data point, SMOTE algorithm selects $k$ nearest neighbors. The neighbors refer to the data points within the minority class that closely resemble the reference point. Synthetic instances are generated by integrating the feature vectors of the reference point and its chosen neighbors.

## 3.3 | Partitioning of the training set

This section presents a novel K-means-like partitioning algorithm that uses a new objective function to detect new patterns of the two a priori classes (fraud and legitimate) of credit card transactions. The proposed method optimizes two criteria: the compactness of the subgroups of the same class and the distances between the representative element (centroid) of one subgroup and the representative elements (centroids) of the subgroups of the second class.

In addition, this objective function uses a weighted Euclidean distance to measure the importance of features in the optimization of the new criterion function.

This new method differs from the conventional K-means algorithm, in which the objective function is based on the internal variability of clusters. Therefore, under these conditions, separation between subgroups is essential to determine various patterns and to detect well-defined subgroups within each original class. The main point is that each subgroup is associated with a distinct weighted Euclidean distance for comparing the elements of the subgroup and its representatives, with the weights that change at each iteration.

Hereafter, as notation, we will use bold uppercase letters to represent matrices and bold lowercase letters to represent vectors.

Let $\mathbf{U} = \begin{pmatrix} \mathbf{U}_F & 0 \\ 0 & \mathbf{U}_L \end{pmatrix}$ be abinary matrix of dimensions $n \times (s_F + s_L)$ used to denote the assignment of the $n$ transactions to one subgroup of the fraud class (respectively legitimate class), where $s_F$ (resp. $s_L$) represents the number of subgroups for fraud (legitimate) class. $\mathbf{U}_F = (u^F_{ip})$ (resp. $\mathbf{U}_L = (u^L_{iq})$) represents the membership matrix for fraud subgroups ($p = 1, \dots, s_F$) (resp. for the legitimate subgroups $q = 1, \dots, s_L$) transactions. In this context, the $i$th row of these matrices, denoted by $\mathbf{u}^F_i(u^F_{i1}, \dots, u^F_{is}, \dots, u^F_{is_F})$ (resp. $\mathbf{u}^L_i = u^L_{i1}, \dots, u^L_{is}, \dots, u^F_{is_L}$)), corresponds to a transaction of the fraud (legitimate) class. Each column $p$ of the matrix $\mathbf{U}_F$ corresponds to the $p$th subgroup, so that when $u^F_{ip} = 1$ means that transaction $i$th of the fraud class belongs to subgroup $p$. On the other hand, if $u^F_{ip} = 0$, means that the transaction is assigned to a different subgroup $p'$. Similarly, each column $q$ of the matrix $\mathbf{U}_L$ corresponds to the subgroup $q$th, of the legitimate subgroup, with $u^L_{iq} = 1$ when the transaction $i$th of the legitimate class belongs to the subgroup $q$ and $u^L_{iq} = 0$, when the transaction is assigned to a different subgroup $q'$.

The centroids of each original class (fraud $F$ and legitimate $L$) are represented by the matrix $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_F \\ \mathbf{Z}_L \end{pmatrix}$ of dimension $(s_F + s_L) \times m$, where $m$ is the number of features, and $\mathbf{Z}_F$ (respectively $\mathbf{Z}_L$) is the subgroup centroids matrix of the fraud (legitimate) class partition. Then, the vector $\mathbf{z}^F_i = (z^F_{i1}, \dots, z^F_{im})$ represents the centroid of the $i$th subgroup of the fraud class partition. $\mathbf{W} = \begin{pmatrix} \mathbf{W}_F \\ \mathbf{W}_L \end{pmatrix}$ is a matrix of weights associated with the subgroup of an a priori class, the vector $\mathbf{w}^F_i = (w^F_{i1}, \dots, w^F_{im})$ is the vector of the $i$th subgroup of the fraud class.

The new objective function combines the within compactness and the between separation criteria, as shown in Equation (3). The proposed technique uses an iterative K-means-like algorithm based on the minimization of the following objective function:

$$\begin{aligned} \Delta(\mathbf{U}, \mathbf{W}, \mathbf{Z}) = &\sum_{j=1}^{m} \left( \sum_{p=1}^{s_F} \frac{w^{\beta}_{pj} \sum_{i=1}^{n} u^F_{ip} (x_{ij} - z^F_{pj})^2}{n_F (z^F_{pj} - z_{Lj})^2} + \sum_{q=1}^{s_L} \frac{w^{\beta}_{qj} \sum_{l=1}^{n} u^L_{lq} (x_{lj} - z^L_{qj})^2}{n_L (z^L_{qj} - z_{Fj})^2} \right) \\ \text{s.t.} \quad & u^F_{ip}, u^L_{lq} \in \{0, 1\}, \ \sum_{p=1}^{s_F} u^F_{ip} = 1 \ \sum_{q=1}^{s_L} u^L_{lq} = 1 \\ & \sum_{j=1}^{m} w^F_{pj} = 1 \qquad \sum_{j=1}^{m} w^L_{pj} = 1, \end{aligned} \tag{3}$$

where:

$\beta$ is a parameter used to tune the weights;

$z_{Fj}$ is the $j$th feature value of the global centroid $\mathbf{Z}_F$ of the group $F$. It is calculated as follows:

$$z_{Fj} = \frac{\sum_{p=1}^{s_F} \sum_{i=1}^{n} u^F_{ip} z^F_{pj}}{\sum_{p=1}^{s_F} \sum_{i=1}^{n} u^F_{ip}}; \tag{4}$$

$z_{Lj}$ is the $j$th feature value of the global centroid $z_L$ of the group $L$. That is calculated as follows:

$$z_{Lj} = \frac{\sum_{q=1}^{s_L} \sum_{i=1}^{n} u^L_{iq} z^L_{qj}}{\sum_{q=1}^{s_L} \sum_{i=1}^{n} u^L_{iq}}. \tag{5}$$

To optimize the objective function, the parameters for the two classes are first initialized. Then, the minimization problem for the partition of the class $F$ is reduced to solving the following equation under constraints:

$$\Delta(\mathbf{U}^F, \mathbf{W}^F, \mathbf{Z}^F) = \sum_{p=1}^{s_F} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{pj}^{\beta} \frac{u_{ip}^F (x_{ij} - z_{pj}^F)^2}{n_F (z_{pj}^F - z_{Lj})^2}$$
$$\text{s.t.} \quad u_{ip}^F \in \{0, 1\}, \quad \sum_{p=1}^{s_F} u_{ip}^F = 1 \tag{6}$$
$$\sum_{j=1}^{m} w_{pj}^F = 1.$$

We can minimize Equation (6) by solving the following problems iteratively:

1. Problem $\Delta 1$: fix $Z^F = \hat{Z}^F$, $W^F = \hat{W}^F$ and solve the reduced problem $\Delta(U^F, \hat{Z}^F, \hat{W}^F)$.
2. Problem $\Delta 2$: fix $U^F = \hat{U}^F$, $W^F = \hat{W}^F$ and solve the reduced problem $\Delta(\hat{U}^F, Z^F, \hat{W}^F)$.
3. Problem $\Delta 3$: fix $U^F = \hat{U}^F$, $Z = \hat{Z}$ and solve the reduced problem $\Delta(\hat{U}, \hat{Z}, W)$.

The problem $\Delta 1$ is solved by

$$u_{ip}^F = \begin{cases} 1 & \text{if } \sum_{j=1}^{m} w_{pj}^{\beta} \frac{(x_{ij} - z_{pj}^F)^2}{n_F (z_{pj}^F - z_{Lj})^2} \leq \sum_{j=1}^{m} w_{rj}^{\beta} \frac{(x_{ij} - z_{rj})^2}{n_L (z_{rj}^F - z_{Lj})^2} \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $1 \leq r \leq s_F$, $r \neq p$.

To solve the problem $\Delta 2$ we derive the gradient of $\Delta$ with respect to $z_{pj}$ as

$$z_{pj}^F = \frac{\sum_{i=1}^{n} u_{ip}^F x_{ij} (x_{ij} - z_{Lj})}{\sum_{i=1}^{n_F} u_{ip}^F (x_{ij} - z_{Lj})}. \tag{8}$$

The problem $\Delta 3$ is solved by setting up a Lagrangian equation $L(W^F, \alpha)$ to $\Delta(\hat{U}^F, \hat{Z}^F, W^F)$ with multiplier $\alpha$.

$$L(W^F, \alpha) = \sum_{p=1}^{s_F} \sum_{j=1}^{m} w_{pj}^{\beta} D_{pj}^F - \alpha \left( \sum_{j=1}^{m} w_{pj} - 1 \right), \tag{9}$$

where

$$D_{pj}^F = \sum_{i=1}^{n} u_{ip} \frac{(x_{ij} - z_{pj}^F)^2}{n_F (z_{pj}^F - z_{Lj})^2}. \tag{10}$$

By equating the derivative of the Equation (9) with respect to both $w_{pj}$ and $\alpha$ to zero, we derive the following equations

$$\frac{\partial L(W^F, \alpha)}{\partial w_{pj}} = \beta w_{pj}^{\beta - 1} D_{pj}^F - \alpha = 0 \Rightarrow w_{pj} = \left( \frac{\alpha}{\beta D_{pj}^F} \right)^{\frac{1}{\beta - 1}}, \tag{11}$$

$$\frac{\partial L(W^F, \alpha)}{\partial \alpha} = -\left( \sum_{j=1}^{m} w_{pj} - 1 \right) = 0 \tag{12}$$

**Algorithm 1.** Clustering algorithm steps for partitioning training set

```
Input : the training set $T = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, shared into fraudulent and legitimate transactions; the optimal number $s_F$ and $s_L$ of fraud and legitimate subgroups respectively;
Output : $\mathbf{U}, \mathbf{Z}, \mathbf{W}$
Randomly initialize $\mathbf{U}^0$, $\mathbf{Z}^0$ and $\mathbf{W}^0$
repeat
Fixed $\hat{\mathbf{U}}^F, \hat{\mathbf{W}}^F, \hat{\mathbf{U}}^L, \hat{\mathbf{W}}^L$ solve the centroids matrix $\mathbf{Z}^F$, $\mathbf{Z}^L$ with Equation (8);
Fixed $\hat{\mathbf{U}}^F, \hat{\mathbf{Z}}^F, \hat{\mathbf{U}}^L, \hat{\mathbf{Z}}^L$ solve the weights matrix $\mathbf{W}^F$, $\mathbf{W}^L$ with Equation (14);
Fixed $\hat{\mathbf{W}}^F, \hat{\mathbf{Z}}^F, \hat{\mathbf{W}}^L, \hat{\mathbf{Z}}^L$ solve the membership matrix $\mathbf{U}^F$, $\mathbf{U}^L$ with Equation (7);
until convergence
```

substituting $w_{pj}$ value from Equation (11) into Equation (12), we have

$$\alpha^{\frac{1}{\beta-1}} = \frac{\beta^{\frac{1}{\beta-1}}}{\sum_{j=1}^{m} \frac{1}{(D^F_{pj})^{\frac{1}{\beta-1}}}} \tag{13}$$

substituting Equation (13) into Equation (11), we have

$$w^F_{pj} = \frac{1}{\sum_{l=1}^{m} \left(\frac{D^F_{pj}}{D^F_{pl}}\right)^{\frac{1}{\beta-1}}}. \tag{14}$$

The incorporation of a weighted distance metric into the clustering algorithm holds promise for improving their performance across numerous aspects. The objective of weighting features is to allocate lower weights to data points showing poorer compactness within subgroups and greater separation between subgroups.

The partition and the estimation of the parameters $\mathbf{U}^L$, $\mathbf{Z}^L$, and $\mathbf{W}^L$ for the group $L$ of legitimate transactions are obtained according to the same optimization procedure.

The clustering algorithm with the new objective function is executed in the following steps (Algorithm 1):

## 3.4 | Forward feature selection based on the clustering weights to improve the classification performance

Feature selection is one of the most widespread and important techniques for decreasing feature dimensionality without compromising the performance of the methodology. Those features do not give the useful information called as irrelevant features and those features do not offer more information called as redundant features. The major goal of the feature selection methods is to reduce the data and the computational complexity and improve the learning performance for accurate prediction.

Multiple methods are present in the ML literature for choosing predictive or non-predictive features. Among these techniques is forward feature selection (FFS), which involves removing non-predictive features in a step-by-step manner. In this research, FFS used to enhance the model performance by picking features that minimize both false legitimate and false fraud, as measured by the F1-score in Equation (23). Consequently, this process improves learning outcomes and reduces training duration.

This study introduces a novel methodology for feature selection. This approach is implemented following the initial phase of clustering, during which the clustering process incorporates a weighted step to assess the significance of variables inside each subgroup. This phase adds to the variable selection process by employing a criterion that is reliant on subgroup weights.

The criterion used to assess the significance of characteristics based on the weight matrix $W$, which comprises weight vectors for each subgroup, is the coefficient of variation (CV). The CV serves as a valuable tool for evaluating the significance of features since it enables a comparative analysis of the dispersion of variables. Additionally, this measure of variability is resilient and not influenced by the scale or unit of measurement used for the data.

**Algorithm 2.** Forward FS based on the clustering weights

```
Input :
$T$: Training set;
$\mathbf{W}$: the weights matrix;
$V_m = \{v_1, \dots, v_m\}$: the set of $m$ features
Output : $V_s = \{v_1, \dots, v_s\}$: Important features
Compute the coefficient of variation vector according to the features of $W$ based on:
    $$CV_{v_j}(W) = \frac{Std(W^{v_j})}{\mu(W^{v_j})} \quad j = \{1, \dots, m\} \qquad (14)$$
    Where $W^{v_j}$ indicate the feature column $v_j$ of the matrix $W$
Sort the features $V_m$ based on the criterion 15 in $Sort_{V_m}$
$V_s = [\ ]$
$V1_0 \leftarrow 0, \quad j \leftarrow 1$
for each $v \in Sort_{V_m}$ do
    Apply the new KNN classifier to $T$ with features $V_s \cup v$;
    Calculate the F1-score of the classifier which is represented by $V1_j$ for $V_s \cup v$;
    if $V1_j > V1_{j-1}$ then
        $V_s \leftarrow V_s \cup v$
    end if
    $V1_j = v$
    $j \leftarrow j + 1$
end for
Return the predictive features $V_s$
```

These studies provide evidence to support the idea that employing feature selection using the FFS method can enhance the performance of classifiers. By finding pertinent features, it becomes possible to enhance the credit card detection model.

The FFS algorithm is executed in the following steps (Algorithm 2):

## 3.5 | K-Fuse algorithm

This research uses a new CCFD model, which incorporates a new supervised classification model. This model is based on an updated version of the KNN algorithm, which exploits the results of clustering and feature selection processes. The goal is to accurately identify the label of new transactions. For this reason, we proposed a Euclidean weighted distance metric, in which weight vectors are automatically determined in the process of searching for subclasses of fraud and legitimate classes within the training set. It is worth noticing that each subgroup is related to its respective weight vector. The main concept of the new variant of the KNN classifier model, which incorporates weighted distance, is that the distance used to compare test transactions to the nearest points in the training set varies for each training transaction. However, it is important to note that transactions belonging to the same subgroup possess identical vector weights. In addition, the CCFD model incorporates the predictive features obtained from the second stage of feature selection, which also exploits the weights from the clustering stage for feature selection. In addition, the new KNN classifier is trained using the updated labels obtained from the K-means method. In addition, the newly generated transactions are assigned to the original classes, which correspond to the legitimate and fraud categories. The procedure is initiated by employing the K-means method to partition each initial class into a distinct set of clusters. The most suitable number of subgroups is identified on the basis of the Silhouette or Rand index. The obtained subgroups serve as a preprocessing step for the KNN technique, allowing it to operate on subsets of data that exhibit improved homogeneity. The integration of this combination enhances the effectiveness and efficiency of the fraud detection model, as it facilitates the model's acquisition of all the necessary patterns within the dataset for its learning procedure.

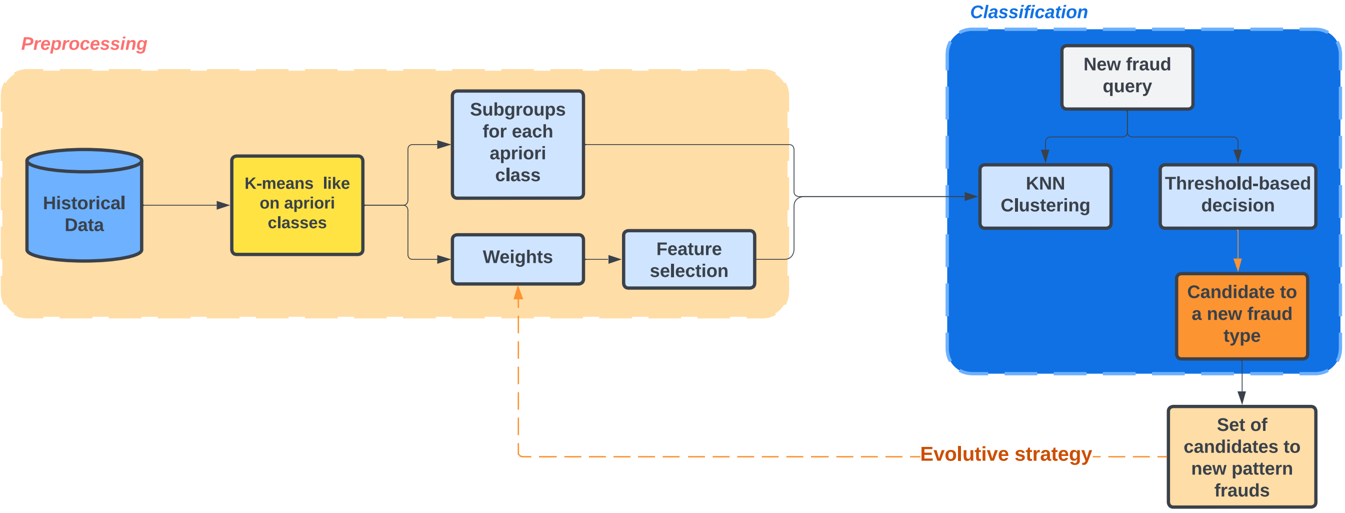

FIGURE 3 Evolving credit card fraud detection model.

After the clustering phase, feature selection is employed to identify and choose the most informative features within the dataset. Fraud detection methodology is employed to categorize new transactions by exploiting newly identified patterns and the most influential variables. The model aims to identify the allocation of a given transaction in a class of fraud or legitime based on KNN using centroids of subgroups as the nearest neighbors. The K-Fuse fraud detection model efficiently integrates the advantages of clustering, feature selection, and classification simultaneously, making it a robust tool for data classification.

Finally, the procedure of CCFD classification consists of three algorithms: a clustering algorithm with a new objective function to partition a priori classes into subgroups; a selection strategy using clustering weights to select the predictive features of the training set classes; a supervised classification based on a new KNN variant classifier that uses the centroids of subgroups as $k$ neighbors (NN) to allocate transactions in their respective classes. $k$ is selected as the optimal value between 1 and $(s_F + s_L)$.

The K-Fuse classifier algorithm for CCFD is performed in the steps 1–14 of Algorithm 3:

### 3.6 | Improving the model adaptability in evolving credit card fraud

Within the constantly evolving field of credit card utilization, fraudsters continually devise novel strategies, posing challenges to our current model. Our K-Fuse model, which exclusively relies on historical data, has limitations in adjusting to these emerging strategies. Drawing inspiration from the concept of zero-shot learning introduced in Kenett,[44] which involves a KNN clustering method, our objective is to expand our technique to better adapt to the dynamic nature of transaction patterns. To achieve this, we plan to integrate a KNN-like technique aimed at improving our model's ability to identify and address new forms of fraud, thereby strengthening our protection against emerging threats.

The early phases of the K-Fuse model involve using a K-means variant to create subgroups within the training dataset and then selecting features based on clustering weights. These steps are considered preprocessing, exclusively employing the training set and can be fully completed before the classification phase. This implies that the practical effectiveness of the fraud detection model may be impacted when new fraud patterns emerge.

These preprocessing tasks are fundamental as they provide the groundwork for the classification process. They involve structuring the data and picking pertinent features, all while ensuring that the algorithm's runtime performance remains unaffected. This is particularly important when our objective is centered around historical data. Our approach to identify new types of fraud transactions progresses through a structured sequence of steps (Figure 3):

1. The first step takes into account the most important features based on their selection in the preprocessing phase, at the query transaction. This involves extracting important features from transactions that can capture the nature of fraudulent behavior, even in cases where the exact model has not been observed.

2. Afterward, we identify the nearest subgroup centroids relative to the new transaction and apply majority voting to detect the original class of this transaction.
3. We focus on the nearest subgroup $C_p$ to the new fraud transaction $\mathbf{x}_i^f$ applying a dynamic thresholding approach to make decisions. Consistently with our methodology where subgroups are characterized by their centroids, a significant deviation of the fraud query from the nearest centroid may indicate a novel type of fraud. To assess this deviation, the distance from the fraud query to the nearest subgroup centroid, denoted as $\mathbf{z}_p^F$ for the unseen transaction $\mathbf{x}_i^f$, is evaluated against a threshold proportional to the maximum distance between the centroid of this subgroup and its constituent items. The threshold is determined by setting a parameter $\gamma > 0$. If the following condition is satisfied:

$$d(\mathbf{x}_i^f, \mathbf{z}_p^F) > (1+\gamma) \max_{x_j \in C_{\ p}} d(x_j, \mathbf{z}_p^F), \tag{16}$$

**Algorithm 3.** K-Fuse combining algorithm with model adaptability

```
Input:
$T = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$: training dataset with fraud and legitimate classes
$s_F,\ s_L$: the optimal number of subgroups within the fraud and legitimate groups respectively
$\mathbf{x'}_l$: new transaction
$k$: number of nearest neighbors centroids to $\mathbf{x'}_l$
$\gamma$: threshold for new potential frauds types
Output: $y_{\mathbf{x'}_l}$ the label of $\mathbf{x'}_l$
Step 1- Dynamic clustering algorithm: partition of the fraud and legitimate classes of $T$ respectively into $s_F$, $s_L$ subgroup using Algorithm 1
return: U, Z and W
Step 2- Feature selection algorithm: select the important features based on FFS Algorithm 2
return: $V_s$ the predictive features
Step 3- KNN classifier using subgroups centroids (representative) as NN
Calculate the distances $d(\mathbf{x'}_l^{V_s}, (\mathbf{z}_p^F)^{V_s})$, $d(\mathbf{x'}_l^{V_s}, (\mathbf{z}_q^L)^{V_s})$ $(p = 1, \dots, s_F \quad and \quad q = 1, \dots, s_L)$.The notation $\mathbf{x'}_l^{V_s}$ denotes the instance $\mathbf{x'}_l$ with exclusively the features from the set $V_s$
Select the subset denoted as $N_k$, which contains the centroids (representatives) representing the $k$ nearest neighbor subgroups of the sample$\mathbf{x'}_l$, this selection is based on the minimal distance between $x'_l$ and subgroups centroids of the both classes
    $N_k(\mathbf{x'}_l) = \{\mathbf{z}_i^g\}_{i=1}^k,$    (17)
    where $g$ represent $F$ for fraud group or $L$ for legitimate group
Allocate $\mathbf{x'_l}$ to its original class, such as:
    $y_{\mathbf{x'}_l} = \begin{cases} 0 & \text{if} \sum_{\mathbf{z} \in N_k(\mathbf{x'}_l)} \mathbb{1}_{\mathbf{Z}^L}(\mathbf{z}) > \sum_{\mathbf{z} \in N_k(\mathbf{x'}_l)} \mathbb{1}_{\mathbf{Z}^F}(\mathbf{z}) \\ 1 & \text{otherwise.} \end{cases}$    (18)
If $y_{\mathbf{x'}_l} = 1$, allocate $\mathbf{x'_l}$ to the emerging fraud patterns by:
    $t_{\mathbf{x'}_l} = \begin{cases} 0 & \text{if } d(\mathbf{x}_i^f, \mathbf{z}_p^F) > (1+\gamma) \max_{x_j \in C_p} d(x_j, \mathbf{z}_p^F) \\ 1 & \text{otherwise.} \end{cases}$    (19)
return: $y_{\mathbf{x'}_l}$
```

$x_i^f$ is not allocated to any of the previous subgroups of frauds and it is assumed as a potential candidate for a new fraud pattern. The K-Fuse including the detection of new fraud types is in Algorithm 3.

4. An extension of the proposed strategy, in an evolutionary perspective, can be considered. When several potential new fraud types are detected, the algorithm is reiterated: subgroups and feature weights are actualized to redefine the new pattern structure for the frauds. To update the weights, we use the Equation (14).

Our methodology combines preprocessing techniques with a KNN-based algorithm to efficiently detect new, unseen patterns of fraud in dynamic credit card transactions. By utilizing information data for internal representation learning and employing clustering weights to select and identify important features, we aim to improve the performance of fraud pattern detection. Additionally, we address the limitations of approximate searches by enhancing how data is represented. This is achieved by training on augmented subgroup versions of the original data groups and employing important feature selection. This integrated method ensures a robust detection mechanism against the constantly evolving challenge of credit card fraud.

## 4 | RESULTS

In this section, we provide a complete overview of the indicators employed to evaluate the performance of CCFD models. The metrics comprise a variety of measurements, such as the confusion matrix, recall, specificity, precision, F-score, and area under the curve (AUC). These metrics collectively contribute to the evaluation of the performance of the fraud detection model. In order to determine the superiority of the suggested strategy K-Fuse, a comparative analysis is performed against established baseline classifiers. The classifiers included in this set are KNN, decision tree (DT), logistic regression (LR), RF, XGBoost, and SVM, these classifiers demonstrated the fastest execution time, as expected given their simplicity, whereas the K-Fuse method, which involves a clustering step to discover hidden patterns in a priori classes, required a slightly longer execution time. This can be explained by the feature selection process, which involves calculating and identifying important features. In this research, the parameter $\beta$ will be set to a fixed value of 2 for both applications.

### 4.1 | Evaluation metrics

For the fraud detection task, we concern about the model's performance to correctly identify the fraud transactions. Therefore, we use recall as one of our metrics. However, if model is not learning effectively and simply predicts all samples as frauds, we also get high recall. To avoid the confusing results, meanwhile, we use precision to measures the accuracy of the positive predictions made by the model, it indicates the proportion of correctly predicted fraud transactions out of all transactions predicted as fraud by the model. We use F1-score as one of metrics. Furthermore, due to the extremely imbalance of sample distribution, we use AUC (insensitive to distribution of samples) as the third metric to evaluate our model more fairly. The AUC is a crucial metric in imbalance classification problems, such as fraud detection, as it indicates the classifier's ability to differentiate between the fraud and non-fraud classes.[45]

The confusion matrix is a tabular representation of the performance of a classification model that helps visualize the model's predictions compared to the actual ground truth. The configuration of the confusion matrix is illustrated in Table 1 and it breaks down the model's predictions into four categories:

- True positive (TP): number of instances in which the model accurately identifies fraudulent transactions as fraudulent.
- True negative (TN): number of legitimate transactions that the model correctly identifies as legitimate.
- False negative (FN): number of fraudulent transactions that the model incorrectly identifies as legitimate.

TABLE 1 Confusion matrix for a binary classification.

| | Legitimate | Fraud |
|---|---|---|
| Legitimate | TN | FP |
| Fraud | FN | TP |

- False positive (FP): number of legitimate transactions that the model erroneously identifies as fraudulent.

The Equation (20) represents the classification recall measurement, it calculates how many positive instances (true labels) are correctly predicted as positive. It is also known as sensitivity or true positive rate. It is formulated as:

$$\text{Recall} = 100 \times \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{20}$$

Precision is a measure that calculates how many positive predictions are correctly identified as positive. It is formulated as follows:

$$\text{Precision} = 100 \times \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{21}$$

Specificity is a measure that calculates how many negative actual labels are correctly identified as negative. The equation below formulates the metric of specificity:

$$\text{Specificity} = 100 \times \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{22}$$

The F1-score is a metric of performance that calculates the weighted harmonic mean of the recall and the precision to provide a single performance measure, taking into account both false positives and false negatives. Formulated as follows:

$$\begin{aligned} \text{F1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{\text{TP}}{\text{TP} + 1/2(\text{FP} + \text{FN})}. \end{aligned} \tag{23}$$

## 4.2 | CD1 dataset results discussion

The problem of imbalanced data, as outlined in Section 3.2, was seen during the course of the studies. The evaluation of our approach's effectiveness was conducted with a focus on data preparation techniques such as feature selection and balancing processes like under-sampling or over-sampling. These techniques are recognized as key steps in enhancing classifier performance, particularly in the domain of fraud detection. Prior to implementing the fraud detection models, we ensured that the training set of our CD1 dataset was balanced by using the SMOTE algorithm for balancing data, as depicted in Figure 4, and the dataset characteristics described in Table 2.

In this discussion, we analyze the experimental results obtained from the implementation of our procedure on the CD1 dataset, which is currently being examined. In order to use K-Fuse, the dataset is partitioned into training and test sets. The training set comprises 70% of the original dataset and is used for training the FFS model. Conversely, the testing set constitutes 30% of the original dataset and is utilized to assess the performance of the model. As a result of the substantial size of the initial dataset CD1, only 6% of the whole dataset was utilized for experimental purposes.

The appropriate number of subgroups for the K-Fuse algorithm was determined by testing different numbers of subgroups for both the legitimate and fraud training classes. The results are presented in Table 3. As documented in the literature, well-established algorithms such as the Rand Index and Silhouette approach can be employed to determine the ideal number of patterns within each class. This research is centered on enhancing the fraud detection model. In order to validate the number of subgroups inside each original class, the recall metric is utilized as a criterion. According to the data presented in the table, it can be observed that the fraud class exhibits an optimal number of three subgroup combinations, while the legitimate group demonstrates four subgroup combinations. The recall metric for these subgroup combinations is calculated to be 93.04%.

We compare K-Fuse to other well-known state-of-the-art CCFD models. This comparison is summarized in Table 4 and illustrated in Figure 5 for the CD1 dataset, which clearly demonstrates the ability of our proposed classifier approach to improve the performance of K-Fuse. In our study, we found that K-Fuse with an FFS that uses a new feature selection criterion and a balancing dataset approach produces the most accurate and precise results in terms of recall, and AUC excluding the precision, F1-score, and specificity in comparison to the other classifier.

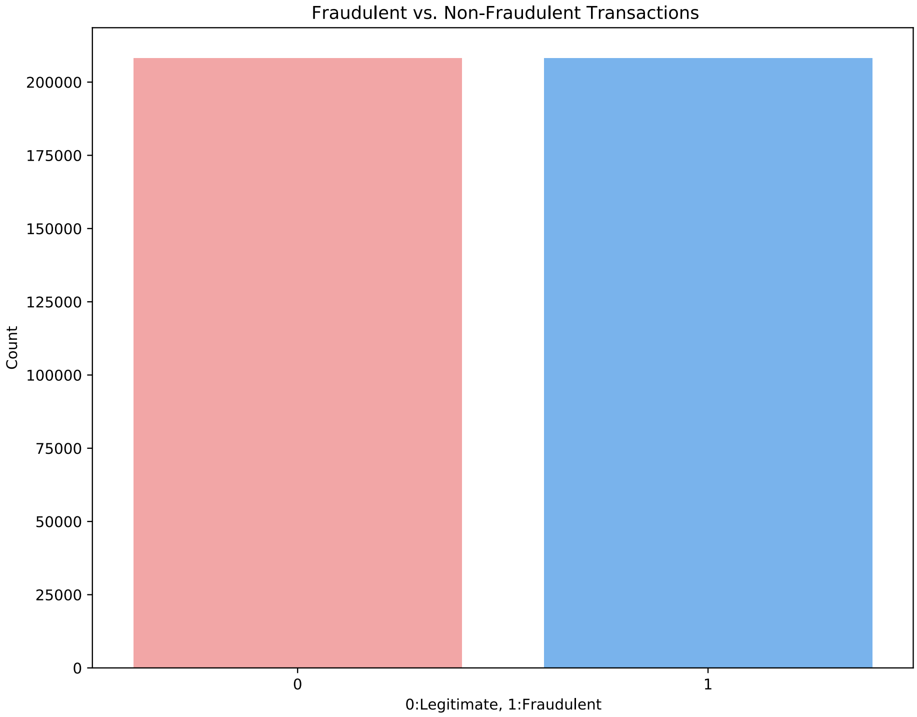

FIGURE 4 Balanced data counts for CD1.

TABLE 2 CD1 dataset characteristics.

| | Legitimate transactions | Fraudulent transactions | Number of features |
|---|---|---|---|
| Original data | 297,418 | 2582 | 75 |
| Training set after SMOTE | 208,190 | 208,190 | 75 |
| Test set | 89,267 | 733 | 75 |

TABLE 3 K-Fuse recall with a distinct number of subgroups for each class.

| Legitimate subgroup's number | 2 | | | 3 | | | 4 | | |
|---|---|---|---|---|---|---|---|---|---|
| Fraud subgroup's number | 2 | 3 | 4 | 2 | 3 | 4 | 2 | 3 | 4 |
| Recall | 76.18 | 77.59 | 77.46 | 77.46 | 78.23 | 93.04 | 69.77 | 74.70 | 76.59 |

TABLE 4 Comparative performance analysis of K-Fuse and other CCFD models on imbalanced data for CD1.

| Classifiers | Recall | Precision | Specificity | F1 score | AUC |
|---|---|---|---|---|---|
| K-Fuse | 93.04 | 5.11 | 85.81 | 9.68 | 89.42 |
| KNN | 40.80 | 22.90 | 98.74 | 29.33 | 69.77 |
| Random forest | 47.06 | 35.78 | 99.22 | 40.65 | 73.14 |
| Decision tree | 42.78 | 17.38 | 98.21 | 24.72 | 70.49 |
| XGBoost | 64.09 | 19.07 | 97.51 | 29.39 | 80.80 |
| Logistic regression | 80.97 | 4.53 | 85.05 | 8.59 | 83.01 |
| SVM | 77.01 | 6.20 | 89.78 | 11.48 | 83.39 |

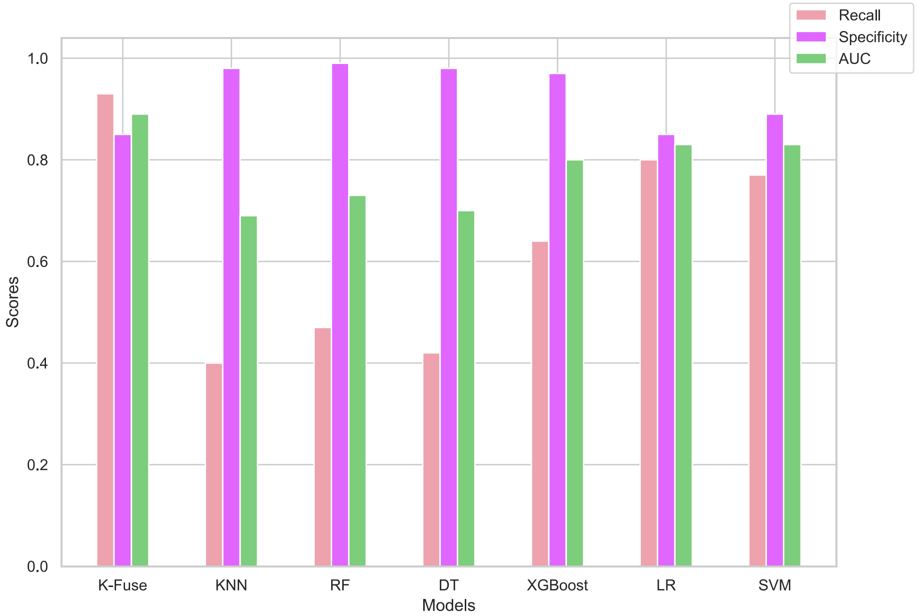

**FIGURE 5** Comparative analysis utilizing different methods for the CD1 dataset.

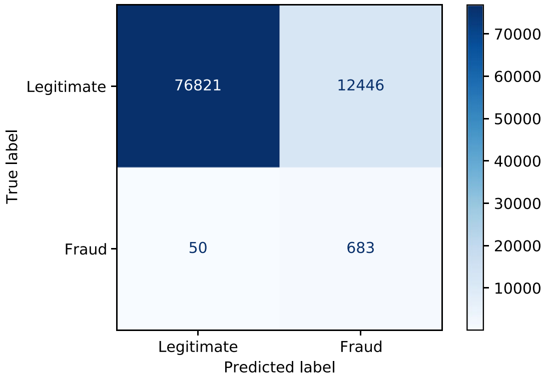

**FIGURE 6** Confusion matrix of CD1 dataset.

This indicates that our model is increasing the true positive rate and decreasing the false negative rate, resulting in enhanced detection of actual instances of fraud. However, it also results in the misclassification of some legitimate transactions as fraudulent, indicating that our model must identify additional patterns within the original group of legitimate transactions. The models included in the comparison with K-Fuse had a significantly greater level of specificity relative to sensitivity. This finding suggests that the models accurately predicted a greater number of legitimate transactions, which constitute the majority class, in comparison to fraudulent transactions, which constitute the minority class.

In the context of credit card fraud detection, similar to other tasks involving imbalanced classification, it holds greater significance to accurately predict the samples belonging to the minority class. Nevertheless, the technique proposed demonstrated exceptional sensitivity, hence highlighting its ability to adapt in accurately forecasting transactions belonging to the minority class.

Furthermore, the confusion matrix of K-Fuse applied to the CD1 dataset in Figure 6 reveals that the instances of fraud that were incorrectly categorized as legitimate, also known as false negatives (FN), amount to 50 out of 733 fraud cases for CD1. The present analysis emphasizes that K-Fuse possesses the capacity to reduce the occurrence of fraudulent transactions. In practical applications, this outcome holds significant value for businesses seeking to enhance their operational efficiency, as it directs their attention towards the mitigation of illegal activities, including fraudulent transactions. The

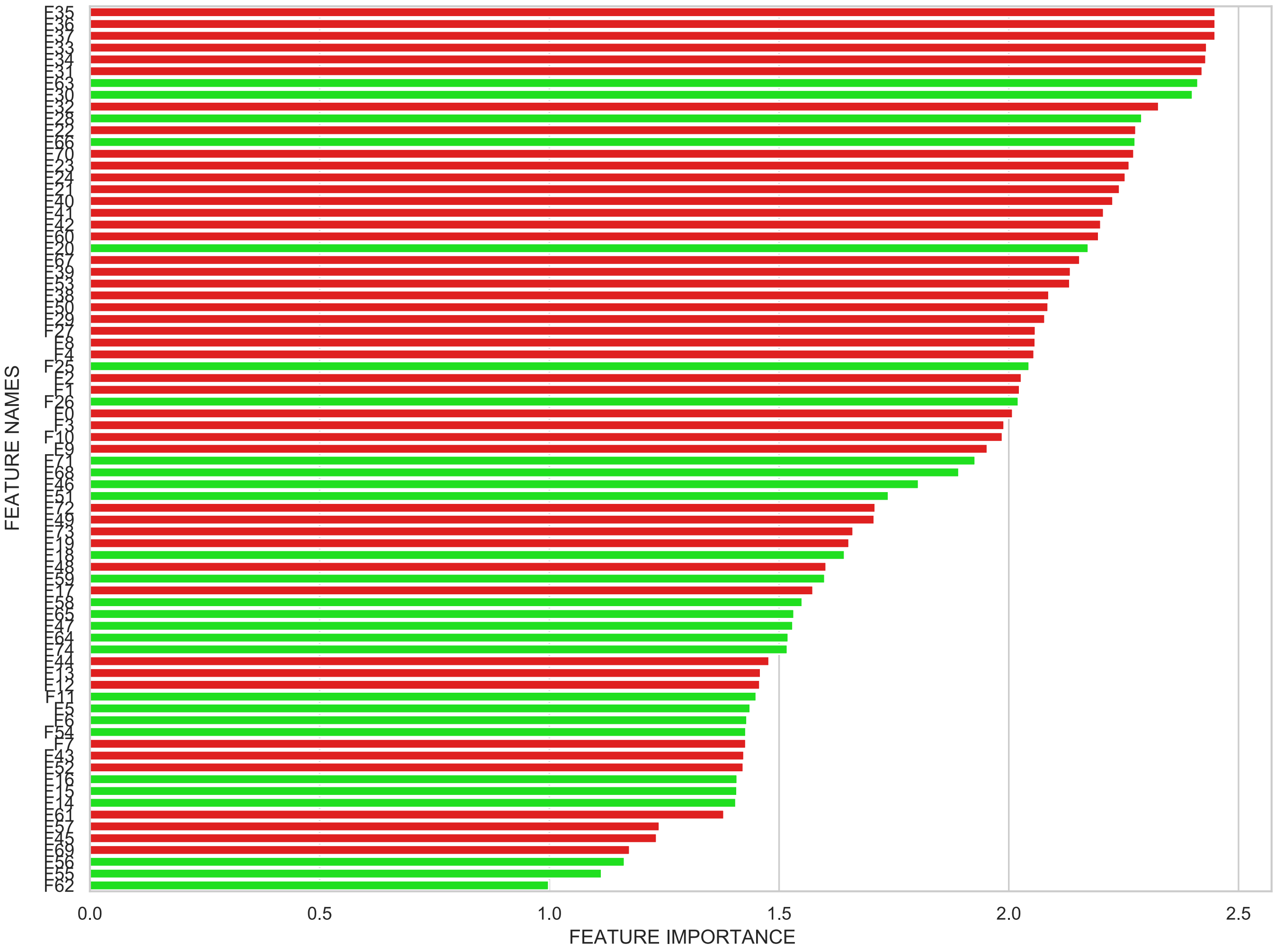

**FIGURE 7** Features importance based on the FFS for CD1.

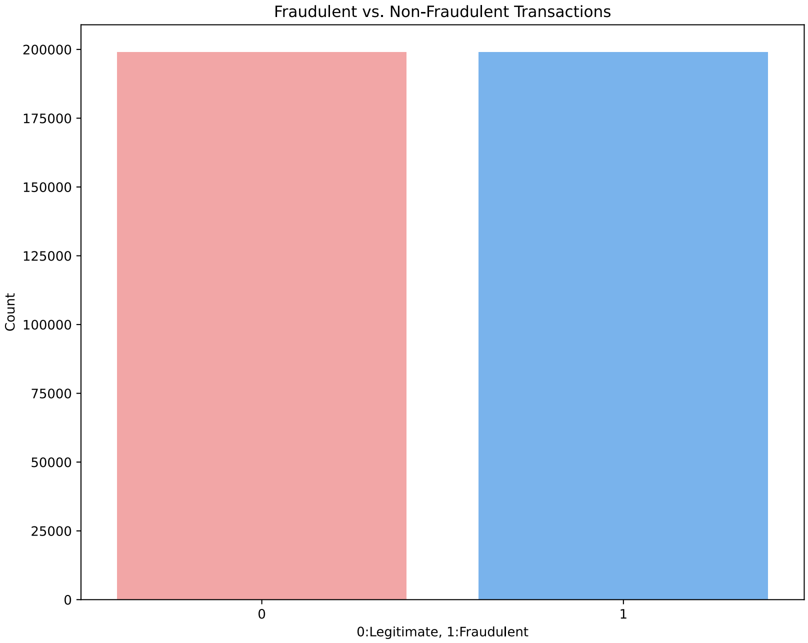

**FIGURE 8** Balanced data counts for CD2.

TABLE 5 CD2 dataset characteristics.

| | Legitimate transactions | Fraudulent transactions | Number of features |
|---|---|---|---|
| Original data | 284,315 | 492 | 30 |
| Training set after SMOTE | 199,013 | 199,013 | 30 |
| Test set | 85,302 | 141 | 30 |

TABLE 6 K-Fuse recall with a distinct number of subgroups for each class for CD2 dataset.

| Legitimate subgroup's number | 2 | | | | | 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Fraud subgroup's number | 2 | 3 | 4 | 5 | 6 | 2 | 3 | 4 | 5 | 6 |
| Recall | 31.20 | 31.91 | 81.56 | 82.97 | 82.97 | 46.09 | 63.82 | 80.85 | 82.26 | 83.68 |

TABLE 7 Comparative performance analysis of K-Fuse and other CCFD models on imbalanced data for CD2.

| Classifiers | Recall | Precision | Specificity | F1 score | AUC |
|---|---|---|---|---|---|
| K-Fuse | 83.68 | 89.39 | 99.98 | 86.44 | 91.83 |
| KNN | 82.97 | 72.22 | 99.94 | 77.22 | 91.46 |
| Random forest | 79.43 | 78.87 | 99.96 | 79.15 | 89.69 |
| Decision tree | 81.56 | 37.70 | 99.77 | 51.56 | 90.66 |
| XGBoost | 80.14 | 84.32 | 99.97 | 82.18 | 90.05 |
| Logistic regression | 64.53 | 79.34 | 99.98 | 73.38 | 82.26 |
| SVM | 70.21 | 97.05 | 99.99 | 81.48 | 85.10 |

dataset's features and the predictive feature sets are depicted in Figure 7. Among them, the 29 features highlighted in green are recognized as the most effective predictors.

After analyzing the CD1 dataset and discussing the findings with our financial partner, who provided us with this dataset, they have expressed interest in our K-Fuse fraud detection approach. This is due to the fact that our model exhibits the highest recall, indicating that it successfully identifies a greater number of fraudulent transactions compared to other models. However, it is important to note that this approach also leads to an increase in the number of false alerts (FP). False alerts occur when the model incorrectly classifies a legitimate transaction as fraudulent. This suggests that the patterns identified by our model may mistakenly categorize certain legitimate transactions as fraudulent. Furthermore, this study represents the original investigation of the dataset, prompting us to propose various ideas. For instance, it is conceivable that the fraud detection software employed by the bank may lack the capability to effectively analyze the complex patterns that fraudsters continually develop. Further preprocessing of the dataset is required prior to training the detection model.

## 4.3 | CD2 dataset results discussion

The proposed K-Fuse methodology is used on CD2 credit card dataset to demonstrate its performance across different scenarios. Following the procedure that was used for the CD1 dataset, we proceeded to split the CD2 dataset into training and test sets. The training set, which accounted for 70% of the original dataset, employs for feature selection with the FFS algorithm before training the K-fuse detection model. Conversely, the testing set, including the remaining 30% of the dataset, was utilized to assess the model's efficacy. It is important to acknowledge that, given the small size of the CD2 dataset, the entirety of the dataset was employed for the purpose of study. In order to achieve balance within the training set of our CD2 dataset prior to the implementation of fraud detection models, the SMOTE method was utilized for the purpose of data balancing, as depicted in Figure 8, and the dataset characteristics described in Table 5.

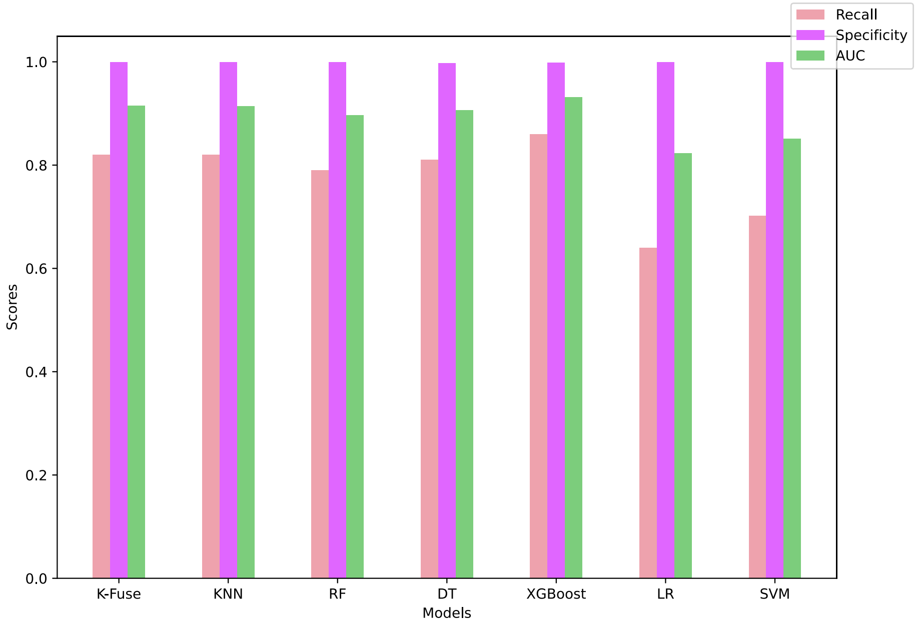

FIGURE 9 Comparative analysis utilizing different methods for the CD2 dataset.

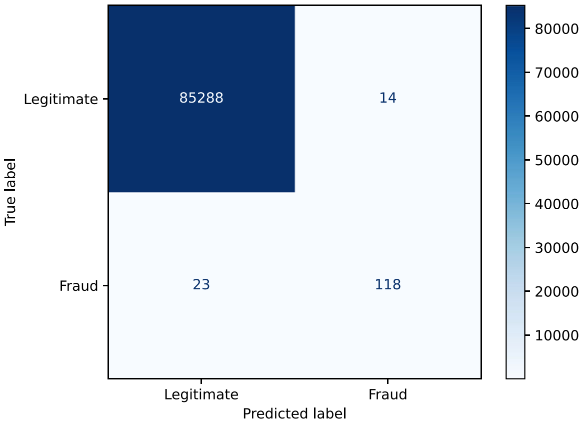

FIGURE 10 Confusion matrix of CD2 dataset.

Based on the findings shown in Table 6, it can be observed that the optimal combination of subgroup numbers for each original class in the training set is associated with the maximum recall. The best combination numbers are three subgroups for the legitimate class, and six subgroups for the fraud class.

The K-Fuse method, as demonstrated in Table 7, attained the following performance metrics on the CD2 dataset: a recall of 83.68%, precision of 89.39%, specificity of 99.98%, F1-score of 86.44%, and an AUC of 91.83%. In the meantime, Figure 9 illustrates the results for recall, specificity, and AUC of the different methodologies. The results indicate that the proposed K-Fuse method exhibited superior performance when compared to existing methods, suggesting its robustness across all metrics performances. In addition, the K-Fuse when used with the CD2 dataset demonstrates a remarkable capacity for identifying fraudulent transactions (with a high recall rate) as well as legitimate ones (with a high specificity rate).

Furthermore, building upon this observed superiority, the K-Fuse methodology introduced in this study incorporates a mechanism of self-learning and self-correction. This design enhances the utilization of fraud pattern information, optimizing its efficacy in detecting hidden instances of fraud within the training set classes, thereby leading to notable enhancements in performance once more.

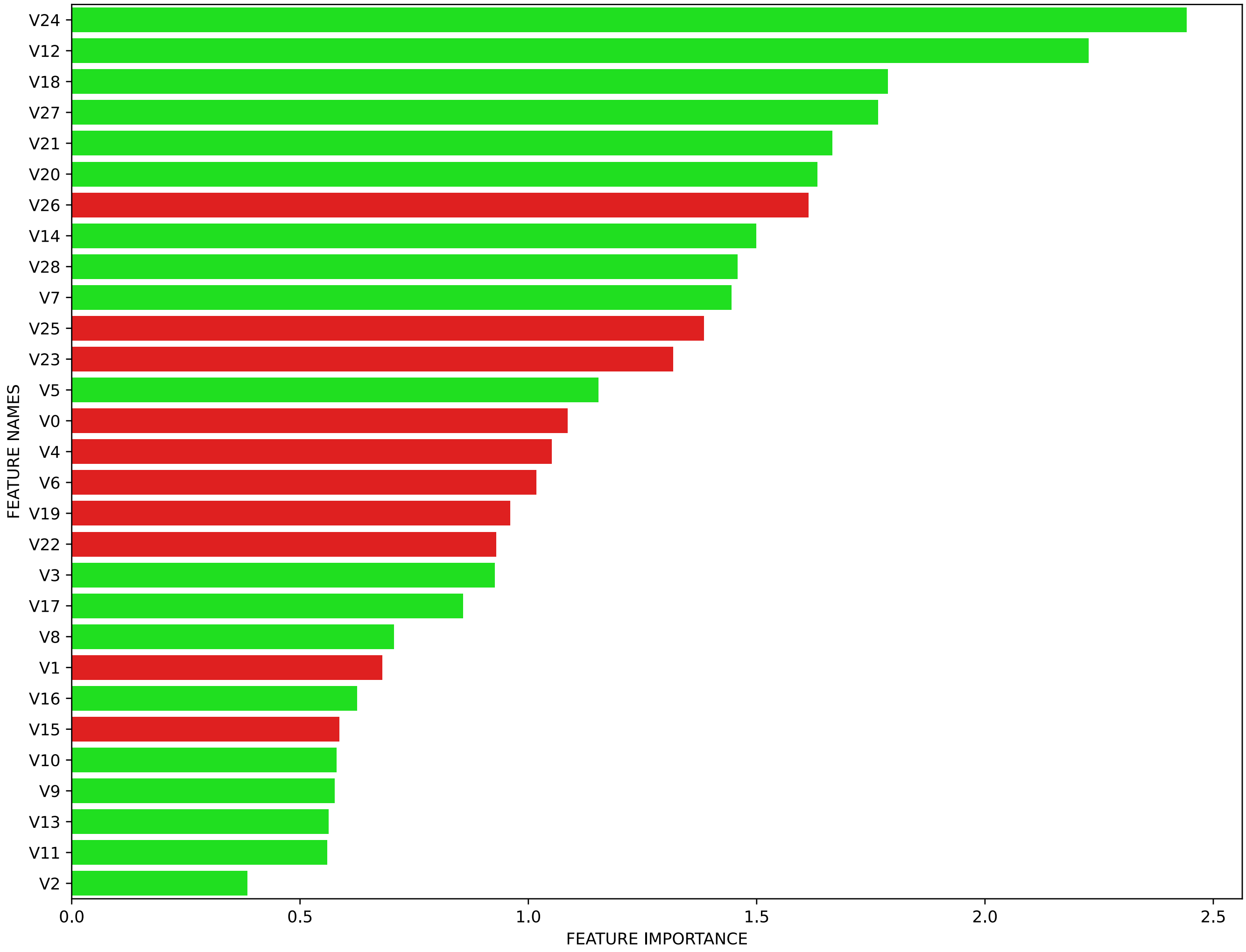

**FIGURE 11** Features importance based on the FFS for CD2.

Moreover, the confusion matrix for the application of K-Fuse on the CD2 dataset, as depicted in Figure 10, indicates that there were 23 instances of fraud that were erroneously classified as legal out of a total of 141 fraud cases for CD2 test set. The current investigation highlights the capability of K-Fuse to mitigate fraudulent transactions and minimize false positive alerts, which refer to legal transactions mistakenly identified as fraudulent by the CCFD model. The properties of the dataset and the sets of predictive features are illustrated in Figure 11. Out of the various qualities, the 19 features that have been highlighted in green are acknowledged as the most efficacious predictors.

## 5 | CONCLUSION AND DISCUSSION

This study introduces the K-Fuse supervised algorithm, an enhanced approach that integrates supervised and unsupervised classification methods for the purpose of detecting fraudulent transactions in the credit card domain. A feature selection phase was employed to choose predictive features based on clustering weights in order to enhance the efficiency of fraud detection. The paradigm we propose comprises three sequential steps: the utilization of clustering with a novel objective function is employed to group both the fraud and legitimate a priori classes to discover hidden patterns that can help to improve the detection model performance. The FFS approach is employed in order to decrease the number of features, utilizing the coefficient of variation criterion. The process of categorizing new transactions is carried out through supervised classification, utilizing the outcomes of the clustering and feature selection stages. In order to ascertain the validity of our findings, we conducted experiments using a dataset provided by our financial collaborator, as well as a widely recognized dataset referenced by the ML community. The results of our model demonstrated its efficacy in identifying a greater number of fraudulent patterns, as measured by five key metrics: recall, precision, specificity, F1 score, and AUC.

The present research employs an unsupervised approach, specifically the K-means algorithm, with a novel objective function. Additionally, a supervised strategy is implemented, namely the KNN algorithm, where the centroids of subgroups are used as the nearest neighbors to assign new transactions. Future research efforts may focus on the development of a time-dependent detection model that employs clustering techniques to categorize transactions occurring in close temporal proximity. This approach will facilitate the identification of transaction patterns within a designated time frame. The performance of a model can be affected by imbalanced data, and one approach to address this issue is to balance the subgroup samples rather than directly manipulating the training set. Furthermore, the utilization of various unsupervised and supervised algorithms can have an influence on the ultimate outcomes of the detection model. Furthermore, our focus will be directed toward tackling the difficulty of identifying new type of fraud patterns unseen in the training phase. We will accomplish this by utilizing transfer learning, which will allow us to enhance the training dataset by including a wide range of simulated fraud patterns and enhance the ability of our model to adapt to new types of fraud without requiring long retraining. By integrating these artificial instances, we expect that our model will possess an improved ability to identify a wider range of fraudulent behaviors, hence enhancing its capabilities to detect fraud in previously unobserved situations. This alternative strategy may potentially lead to changes in the model findings.

By incorporating these advanced ML methods into our fraud detection framework, our goal is to not only improve the model's performance to adjust to developing fraud trends but also to contribute to the advancement of more robust, efficient, and scalable fraud detection systems. This study will facilitate future research on the utilization of advanced ML techniques to prevent financial fraud, with substantial consequences for both the academic community and the financial industry as a whole.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ACKNOWLEDGMENTS

The authors would thank the editor, the associate editor, and the referees for their valuable comments and constructive suggestions.

## ENDNOTE

*https://data.world/raghu543/credit-card-fraud-data.

## REFERENCES

1. Barker KJ, D'amato J, Sheridon P. Credit card fraud: awareness and prevention. *J Financ Crime*. 2008;15:398-410.
2. Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G. Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE Trans Neural Netw Learn Syst*. 2017;29:3784-3797.
3. Li H, Wei M. Fuzzy clustering based on feature weights for multivariate time series. *Knowl-Based Syst*. 2020;197:105907.
4. Kumar MS, Soundarya V, Kavitha S, Keerthika E, Aswini E. Credit card fraud detection using random forest algorithm. *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*. IEEE; 2019:149-153.
5. Liu C, Chan Y, Alam Kazmi SH, Fu H. Financial fraud detection model: based on random forest. *Int J Econ Financ*. 2015;7:178-188.
6. Fu K, Cheng D, Tu Y, Zhang L. Credit card fraud detection using convolutional neural networks. In: Hirose A, Ozawa S, Doya K, Ikeda K, Lee M, Liu D, eds. *Neural Information Processing*. Springer; 2016:483-490.
7. Patidar R, Sharma L. Credit card fraud detection using neural network. *Int J Soft Comput Eng*. 2011;1:32-38.
8. Itoo F, Meenakshi, Singh S. Comparison and analysis of logistic regression, naïve bayes and KNN machine learning algorithms for credit card fraud detection. *Int J Inf Technol*. 2021;13:1503-1511.
9. Sahin Y, Duman E. Detecting credit card fraud by ANN and logistic regression. *2011 International Symposium on Innovations in Intelligent Systems and Applications*. IEEE; 2011:315-319.
10. Gyamfi NK, Abdulai J-D. Bank fraud detection using support vector machine. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE; 2018:37-41.
11. Zhang D, Bhandari B, Black D, et al. Credit card fraud detection using weighted support vector machine. *Appl Math*. 2020;11:1275.
12. Adepoju O, Wosowei J, Jaiman H, et al. Comparative evaluation of credit card fraud detection using machine learning techniques. *2019 Global Conference for Advancement in Technology (GCAT)*. IEEE; 2019:1-6.
13. Ganji VR, Mannem SNP. Credit card fraud detection using anti-k nearest neighbor algorithm. *Int J Comput Sci Eng*. 2012;4:1035-1039.
14. Asha R, Suresh Kumar KR. Credit card fraud detection using artificial neural network. *Glob Transit Proc*. 2021;2:35-41.

15. Xuan S, Liu G, Li Z, Zheng L, Wang S, Jiang C. Random forest for credit card fraud detection. *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE; 2018:1-6.
16. Chougule P, Thakare A, Kale P, Gole M, Nanekar P. Genetic k-means algorithm for credit card fraud detection. *Int J Comput Sci Inf Technol*. 2015;6:1724-1727.
17. Zaslavsky V, Strizhak A. Credit card fraud detection using self-organizing maps. *Inf Secur*. 2006;18:48-63.
18. Rtayli N, Enneya N. Selection features and support vector machine for credit card risk identification. *Procedia Manuf*. 2020;46:941-948.
19. Shamsudin H, Yusof UK, Jayalakshmi A, Khalid MNA. Combining oversampling and undersampling techniques for imbalanced classification: a comparative study using credit card fraudulent transaction dataset. *2020 IEEE 16th International Conference on Control & Automation (ICCA)*. IEEE; 2020:803-808.
20. Alam TM, Shaukat K, Hameed IA, et al. An investigation of credit card default prediction in the imbalanced datasets. *IEEE Access*. 2020;8:201173-201198.
21. Singh A, Ranjan RK, Tiwari A. Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms. *J Exp Theor Artif Int*. 2022;34:571-598.
22. Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. *Phil Trans R Soc A*. 2009;367:4237-4253.
23. Sirimongkolkasem T, Drikvandi R. On regularisation methods for analysis of high dimensional data. *Ann Data Sci*. 2019;6:737-763.
24. Al-Yaseen WL, Idrees AK, Almasoudy FH. Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system. *Pattern Recognit*. 2022;132:108912.
25. Maldonado S, Weber R. A wrapper method for feature selection using support vector machines. *Inform Sci*. 2009;179:2208-2217.
26. Ravisankar P, Ravi V, Rao GR, Bose I. Detection of financial statement fraud and feature selection using data mining techniques. *Decis Support Syst*. 2011;50:491-500.
27. Wang H, Zhu P, Zou X, Qin S. An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering. *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)*. IEEE; 2018:94-98.
28. Carcillo F, Le Borgne Y-A, Caelen O, Kessaci Y, Oblé F, Bontempi G. Combining unsupervised and supervised learning in credit card fraud detection. *Inform Sci*. 2021;557:317-331.
29. Eshghi A, Kargari M. Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty. *Expert Syst Appl*. 2019;121:382-392.
30. Alam TM, Shaukat K, Khan WA, et al. An efficient deep learning-based skin cancer classifier for an imbalanced dataset. *Diagnostics*. 2022;12:2115.
31. Dittman DJ, Khoshgoftaar TM, Napolitano A. The effect of data sampling when using random forest on imbalanced bioinformatics data. *2015 IEEE International Conference on Information Reuse and Integration*. IEEE; 2015:457-463.
32. Makki S, Assaghir Z, Taher Y, Haque R, Hacid M-S, Zeineddine H. An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*. 2019;7:93010-93022.
33. Munkhdalai T, Namsrai O-E, Ryu KH. Self-training in significance space of support vectors for imbalanced biomedical event data. *BMC Bioinformatics*. 2015;16:1-8.
34. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell*. 2016;5:221-232.
35. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21:1263-1284.
36. Luque A, Carrasco A, Martín A, de Las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit*. 2019;91:216-231.
37. Dal Pozzolo A, Caelen O, Bontempi G. When is undersampling effective in unbalanced classification tasks? In: Appice A, Rodrigues P, Santos Costa V, Soares C, Gama J, Jorge A, eds. *Machine Learning and Knowledge Discovery in Databases*. Springer; 2015:200-215.
38. Kotsiantis S, Kanellopoulos D, Pintelas P, et al. Handling imbalanced datasets: a review. *GESTS Int Trans Comput Sci Eng*. 2006;30:25-36.
39. Ramyachitra D, Manikandan P. Imbalanced dataset classification and solutions: a review. *Int J Comput Bus Res*. 2014;5:1-29.
40. Hancock J, Khoshgoftaar TM, Johnson JM. The effects of random undersampling for big data medicare fraud detection. *2022 IEEE International Conference on Service-Oriented System Engineering (SOSE)*. IEEE; 2022:141-146.
41. Salekshahrezaee Z, Leevy JL, Khoshgoftaar TM. The effect of feature extraction and data sampling on credit card fraud detection. *J Big Data*. 2023;10:1-17.
42. Zhang F, Liu G, Li Z, Yan C, Jiang C. GMM-based undersampling and its application for credit card fraud detection. *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE; 2019:1-8.
43. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357.
44. Kenett R. Zero-shot learning. *Wiley StatsRef: Statistics Reference Online*. Wiley; 2023.
45. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowl Manag Process*. 2015;5:1-11.

**How to cite this article:** Sabri M, Verde R, Balzanella A. K-Fuse: Credit card fraud detection based on a classification method with a priori class partitioning and a novel feature selection strategy. *Appl Stochastic Models Bus Ind*. 2024;40(4):1060-1081. doi: 10.1002/asmb.2868