Hindawi
Security and Communication Networks
Volume 2021, Article ID 9293877, 8 pages
https://doi.org/10.1155/2021/9293877

Research Article

Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques

Abolfazl Mehbodniya,[1] **Izhar Alam**,[2] **Sagar Pande**,[2] **Rahul Neware**,[3] **Kantilal Pitambar Rane**,[4] **Mohammad Shabaz**,[5,6] **and Mangena Venu Madhavan**[2]

[1]*Kuwait College of Science and Technology (KCST), Doha, Area, 7th Ring Road, Kuwait*
[2]*School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*
[3]*Department of Computing, Mathematics and Physics, Høgskulen på Vestlandet, Bergen, Norway*
[4]*KCEs COEM Jalgaon, Maharashtra, India*
[5]*Arba Minch University, Arba Minch, Ethiopia*
[6]*Department of Computer Science and Engineering, Chandigarh University, Ajitgarh, India*

Correspondence should be addressed to Mohammad Shabaz; mohammad.shabaz@amu.edu.et

Received 28 July 2021; Revised 21 August 2021; Accepted 23 August 2021; Published 11 September 2021

Academic Editor: Chinmay Chakraborty

Healthcare sector is one of the prominent sectors in which a lot of data can be collected not only in terms of health but also in terms of finances. Major frauds happen in the healthcare sector due to the utilization of credit cards as the continuous enhancement of electronic payments, and credit card fraud monitoring has been a challenge in terms of financial condition to the different service providers. Hence, continuous enhancement is necessary for the system for detecting frauds. Various fraud scenarios happen continuously, which has a massive impact on financial losses. Many technologies such as phishing or virus-like Trojans are mostly used to collect sensitive information about credit cards and their owner details. Therefore, efficient technology should be there for identifying the different types of fraudulent conduct in credit cards. In this paper, various machine learning and deep learning approaches are used for detecting frauds in credit cards and different algorithms such as Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, and the Sequential Convolutional Neural Network are skewed for training the other standard and abnormal features of transactions for detecting the frauds in credit cards. For evaluating the accuracy of the model, publicly available data are used. The different algorithm results visualized the accuracy as 96.1%, 94.8%, 95.89%, 97.58%, and 92.3%, corresponding to various methodologies such as Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, and the Sequential Convolutional Neural Network, respectively. The comparative analysis visualized that the KNN algorithm generates better results than other approaches.

1. Introduction

The popularization of credit cards is across many fields, and healthcare is one among them. Because of credit cards, the online transaction has become more convenient and more accessible. However, fraud transaction impacts the massive loss of capital every year which might increase in the coming year. The system for detecting the fraud might be composed of a manual process and the expertise algorithm for detecting the fraud automatically. The automatic operation can be based upon all previous ways of fraud transactions happened. The manual method is estimated by different fraud investigators who check the separate transaction and generate binary feedback on every transaction. Fraud cases in the transaction are the primary barrier while enhancing the e-commerce and also cause a massive loss in the economy. Hence, detection of fraud is essential while doing transactions in an online environment.

Detection of fraud is the process of analyzing the behavior of card holders' transactions to know whether the conducted transaction is genuine. Frauds in credit cards signify the illegal use of information in credit cards and

completing a transaction. While transacting physically, the involvement of credit card is there while the digital transaction is conducted utilizing the Internet or a telephone as information such as card number, its verification number, and its expiry date is collected through different means. Commonly, two different methodologies are conducted for anomaly detection in a transaction that has been conducted digitally. First, classification is used for determining whether the conducted transaction is genuine or fraudulent. Such approaches help identify the aforementioned types of conducted fraud, which helps construct the different models based upon all earlier patterns of fraud. Detection of the anomaly was conducted by the comparative analysis of data based upon the historical data of the transaction and the newly conducted transaction. It helps to identify all the possible potential of fraud transaction as fraud transaction shows deviation in its behavior from the average transaction. However, detection of fraud through anomaly requires a massive amount of successive data of different behaviors of average transaction of that cardholder. Different frauds in a credit card can be categorized as fraud in external card or inner card. Fraud in inner card happens due to commitment of false identity between the bank and the cardholders, and fraud in external card includes the usage of a stolen credit card to withdraw the cash by dubious means. However, different expertises use different computational methodologies for detecting the frauds in credit cards. Credit card fraud detection is associated with many challenges, such as dynamic or the fraudulent behavior of credit cardholders. Such kinds of activities can be identified using the popular technology called artificial intelligence through machine learning and deep learning algorithms. In particular, in this scenario, one needs to identify whether the cardholder is genuine or fraudulent, i.e., classifying the cardholders. Classification of related applications can be made through some of the ML (machine learning) algorithms such as KNN (K-Nearest Neighbor), Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, and Neural Networks.

This paper consists of comparative analysis conducted between sequential convolutional neural networks and the many machine learning algorithms such as KNN (K-Nearest Neighbor), Random Forest, Decision Tree, Logistic Regression, and Naive Bayes. This paper enhanced the handling of the massive amount of imbalanced data collected from different frauds happened in credit cards, and the dataset is publicly available. In this paper, the main contribution can be summarized as dealing with the different problems related to fraud detection with the help of different machine learning approaches, and at last, from the obtained result, some suggestions and the future work related to detecting the fraud in credit cards are presented.

2. Related Work

This section reviews different fraud detection technologies with a sequential model and the different machine learning approaches. Many credit card financial applications with their transaction history are reviewed. Classification of different transactions related to a credit card mainly falls under the problem of binary classification as it will be a legitimate transaction (false class) or a genuine transaction (true class).

Awoyemi et al. in 2017 [1] investigated severely distorted credit card fraudulent information; this research analyzes the efficiencies of various methodologies such as Naive Bayes, KNN, and Logistic Regression. Credit card transaction information-based data including 284,807 transactions were gathered from European customers. On the distorted information, a combination strategy of undersampling and oversampling is used. The original and preprocessed data are subjected to three procedures. Python is used to carry out the task. The findings reveal that Naive Bayes, K-Nearest Neighbor, and Logistic Regression classifiers have an optimum accuracy of 97.92%, 97.69%, and 54.86%, respectively. KNN outperforms Naive Bayes and Logistic Regression methods, according to the comparison findings. Dal Pozzolo et al. in 2017 [2] proposed three key contributions. First, with the aid of their research assistance, the authors offer a formalization of the fraud-identification issue that accurately reflects the working circumstances of FDSs that monitor enormous flows of credit card transactions daily. The authors also showed how to utilize the most relevant evaluation metrics for fraud identification. Second, the authors devised and tested a unique learning approach for dealing with class imbalance, idea drift, and verification delay. Third, the authors illustrated the influence of class imbalance and idea drift in a real-world information stream with more than 75 million transactions permitted over three years in their studies. To train the behavior characteristics of regular and anomalous transactions, two types of random forests are employed. The framework proposed by Xuan et al. in 2018 [3] compared and analyzed the effectiveness of various random forests with various classification models in terms of credit fraud identification. The data for these tests came from a Chinese e-commerce firm. To include transactional sequences, Jurgovsky et al. in 2018 [4] framed the fraud identification issue as a sequence classification job in their article and used long short-term memory networks. In addition, the system incorporates cutting-edge attribute aggregation techniques and reports the framework findings using standard retrieval measures. The LSTM increases identification accuracy on offline transactions where the cardholder is present physically at merchants when compared to a benchmark Random Forest classifier. Manual attribute aggregation techniques are beneficial to both sequential and nonsequential learning systems. Following a review of true positives, it was discovered that both techniques tend to detect distinct types of frauds, indicating that the two should be used together.

The study by Varmedja et al. in 2019 [5] demonstrated several methods for identifying transactions as fraudulent or legitimate. The dataset utilized in the study was the credit card fraud identification dataset. The SMOTE method was employed to oversample the dataset since it was highly unbalanced. In addition, attributes were chosen and the dataset was divided into two fragments: training data and test data. Logistic Regression, Random Forest, Naive Bayes, and Multilayer Perceptron were the technologies utilized in

the research. The study demonstrates that each technology is capable of detecting credit card fraud with high accuracy. The developed framework may be used to identify additional anomalies. Credit card fraud identification systems that utilize supervised learning methodologies rely on the idea that fraudulent tendencies may be learned from an examination of prior transactions.

Nevertheless, the process gets complicated when it must account for modifications in customer behavior and criminals' capacity to develop new fraud patterns. Unsupervised learning methodologies can aid fraud identification models in detecting abnormalities in this situation. Carcillo et al. in 2019 [6] offered a hybrid approach for improving fraud identification accuracy by combining supervised and unsupervised methodologies. On a real, labeled credit card fraud identification dataset, unsupervised anomaly ratings generated at various degrees of granularity are analyzed and evaluated. The combination is effective, as evidenced by experimental findings, and improves identification accuracy. Machine learning techniques are employed to identify credit card fraud in the research proposed by Randhawa et al. in 2018 [7]. First, conventional methodologies are utilized. After that, hybrid approaches based on AdaBoost and popular voting are utilized. A publicly accessible credit card dataset is utilized to test the framework's effectiveness. The data are then evaluated using a real-time credit card dataset from a financial organization. In addition, distortion is introduced into the data samples to test the techniques' resilience. The experimental findings show that the popular voting approach detects credit card fraud instances with a high degree of accuracy.

De Sá et al. in 2018 [8] proposed the Fraud-BNC methodology to identify credit card fraud issues. The proposed methodology is based on the Bayesian network classification model. Fraud-BNC was created naturally using a dataset from PagSeguro, Brazil's most prominent online payment platform, and evaluated alongside two cost-sensitive categorization methods. The acquired findings were compared to seven other methodologies and assessed for the data classification issue and the methodology's financial efficiency. Fraud-BNC emerged as the most robust methodology for achieving a good balance between the two points of view, increasing the existing organization's financial efficiency by up to 72.64%. A credit card fraud identification model was created by Sailusha et al. in 2020 [9] to identify fraudulent actions. The goal of this research is to concentrate on machine learning methodologies. The Random Forest methodology and the AdaBoost methodology were utilized. The two methodologies' accuracy, precision, recall, and $F1$-score are utilized to compare their outcomes. The confusion matrix is utilized to generate the ROC curve. The performance metrics such as accuracy, precision, recall, and $F1$-score of these two methodologies were compared. The methodology with the best performance metrics is deemed the best methodology for identifying fraud.

Economic fraud has proven to be a threat and has had a significant influence on the financial system. Data mining is one of the approaches that has proven effective in identifying credit card fraudulence in online transactions. Credit card fraudulent identification has proven difficult due to two issues: the characteristics of fraudulent and regular behavior vary over time and the datasets utilized are highly biased. The framework proposed by Bagga et al. in 2020 [10] intended to compare the efficiency of various methodologies such as Logistic Regression, Naive Bayes, Random Forest, KNN, AdaBoost, Multilayer Perceptron, Pipelining, and Ensemble Learning on the information of credit card fraudulence. The variables utilized and the approach employed to identify fraud impact the effectiveness of fraud identification.

2.1. Credit Card Fraud. The comprehensive analysis of different technologies related to fraud detection is essential while solving the different problems related to detecting fraud in credit cards. The most popular algorithm for detecting frauds in credit cards is inspired by nature. Application fraud relates to the criminal who owns a credit card from different issuing companies by spreading false data related to the cardholder [11]. In behavior frauds, the criminal thieves the detail related to the account and the password related to that account and uses that for withdrawing the money. Credit card fraud is more accessible as more money can be earned with less amount of risk in less duration of time.

Recently, many commercial banks adopted the method of fraud detection based upon the behavior related to the cardholder. Mostly, the fraud detection works upon the cardholder behavior pattern of using the card and relates all the transactions based upon the pattern for detecting the unknown transaction [12]. The sequence pattern of credit card transactions mainly relates to the Hidden Markov Model (HMM), which identifies the effectiveness based on credit card fraud [13]. Initially, the HMM is trained with a typical pattern of transaction of the cardholder. Then, when a so ever transaction happens, the new transaction is compared with the pattern of the trained model and if it is denied by the HMM [14], then it signifies that there is a fraud transaction. There is also a two-level sequence alignment technique where both anomaly detection and misuse sequence detection are combined. Here, in these models, profile analyzers were implemented for analyzing and determining the typical pattern between all the transaction sequences related to the cardholder with the past sequence of transactions. Then, the profile analyzer detects the unusual transaction that happened based upon the past and possible transactions and finally states whether the happened transaction is genuine or fraud. Most of the applications related to e-commerce use the signature-based technique for making the deviation related to user's behavior and consequently generalized all the potential behaviors of the fraud [13]. However, mostly they rely upon the clickstream of the signature which used multiple features of the transaction as it generates better results than a single transaction feature. The aggregating profile method exploits the pattern inherent concerning the transaction in time series which detects the fraud of all the transactions online at the end of a particular duration [15]. Here, they evaluate the data with various

techniques such as Random Forest, Logistic Regression, and Support Vector Machine to predict the different frauds related to the credit card with the aggregation technique. However, this aggregation method fails to detect the real-time fraud that happens in the transaction with the credit card.

2.2. Feature Selection. The fraud detection system basis relies upon the behavior analysis of the cardholder. The profile of total expenditure is analyzed with help of optimal variable selection which focuses on the unique behavior of the transaction done through credit cards. The variable compares the current transaction with all the past transactions through which it has been trained. It falls under the following five different variable types: statistics of all transactions, merchant statistics, regional statistics, number of transactions, and the statistics related to the amount of transaction. Thus, through optimal variable selection, both the legitimate and the fraudulent profile can be separated easily, which helps distinguish between the transactions and enhances the system for detecting the frauds related to a credit card [15].

Currently, payment through both online and offline modes has become more common using the credit card. Hence, the rate of fraud accelerates, which brings a massive loss for financial and e-commerce companies. Fraud detection through the traditional method consumes a lot of time; thus, it needed some artificial intelligence models for detecting and tracking out the fraud in credit cards [16]. These techniques of intelligence hold many techniques based upon computational intelligence. The fraud detection system is based upon the supervised and the unsupervised learning method. The fraud detection through the supervised technique depends on the transaction based on fraudulent and legitimate and then newly occurred transaction classified based upon the learned model, while in an unsupervised model of fraud detection, the transactions that lie in outliers are the mainly considered transactions related to the fraud. The algorithms such as backpropagation of error signal with its forward pass and backward pass are implemented for fraud detection [17].

2.3. Comparative Studies. It includes the study of different related issues associated while detecting the frauds in credit cards. Comparative studies help in investigating different credit card-related fraud and nonfraud-related transactions, leading to better accuracy with great learning of an algorithm. The result visualized from the original dataset which leads towards training data is balanced with the help of a metalearning classifier which enhanced the performance of the model. Naive Bayes and Logistic Regression are compared in [18].

In minimum cases, it was observed that the performance of Logistic Regression is less than that of Naive Bayes. However, such a scenario is observed in data with fewer attributes and a small dataset [19]. Three classification methods (Logistic Regression, Decision Tree, and Neural Networks) are compared and tested for fraud detection applicability. The result visualized that the Neural Network classification technique generates better results compared to two other algorithms [20]. The Bayesian learning and the theory of Dempster–Shafer are fused for investigating frauds in credit cards, and it is observed that it has nearly 5% of false-positive rate [21]. Support Vector Machines with Decision Tree are investigated for detecting the fraud, and the result visualized that the classifier of Decision Tree outperforms SVM approaches [22]. The performance of Logistic Regression is evaluated with different approaches of data mining such as Random Forest and Support Vector Machine, and the result visualized that the performance of Logistic Regression is in undersampling level while the performance of the SVM trends to enhance in training data with the lower proportion. In paper [23], there is a comparison between the different classification models such as Logistic Regression and different artificial neural networks are developed to train and test on a dataset of fraud detection with highly skewed data [22]. Its results visualized that the artificial neural network performs better than the Logistic Regression for investigating the fraud related to the credit card. Classifier with Logistic Regression overfits the data while training due to insufficient data, which is a significant issue that causes a fall in its accuracy [24].

The classifier techniques such as Naive Bayes, Neural Network, and Decision Tree are trained and tested. It was observed that a huge database classifier such as Neural Network generates better results than another algorithm [25]. However, usually training and testing the huge dataset with a Neural Network consume a lot of time. Classifiers such as Bayesian take minimum time for training, but it is suitable for the lower or average data size [26]. The problem related to both classification and regression can be solved using a Support Vector Machine by arranging the sample to the category or many classifiers of binary-linear that consist of the nonprobabilistic sample [27]. In the probabilistic sample, the HMM is mainly used for representing different models of classification and regression. In sequential data, the HMM is used for learning succession patterns in abnormal and standard data. The likelihood transaction is used to generate a score for detecting the anomaly [28]. The recurrent neural network (RNN) lies in a nonprobabilistic model. Discriminatively, the RNNs are trained to predict the label of transactions and later generate the sequence of transactions for detecting fraud in credit cards [29]. Scalar variables are linked with Linear Regression by locating the observed data in the linear equation modeled by the function of linear predictor and unknown parameters calculated from the fraud detection data [11,30]. The summary of different machine learning techniques and their limitations is given in Table 1.

3. Experimental Setup and Methods

This section explains using a dataset in the experiment and different deep learning and machine learning classifiers such as Logistic Regression, Naive Bayes, Decision Tree, KNN, and the sequential model. All these algorithms perform different stages before generating the classifier such as data

Table1: Limitations of machine learning techniques.

Model	Strength	Limitations
Bayesian	Provide better results in problems of binary classification and suitable for analyzing the real-time data	Required better detection related to the abnormal and expected behavior of fraud cases
Neural Network	Suitable for problems related to binary classification, mostly used for detecting the fraud	Required huge computation, can be denied for real-time operation, and retraining is essential in terms of newly arrived fraud cases
Decision Tree	Implementation is more straightforward with low power of computation and suitable for analyzing the real-time data	Overfitting may rise if the information of the underlying domain does not set in training data
Logistic Regression	Implementation is easy and fraud detection is based on historical data	Performance of classification is lacking when compared with methods of data mining
Linear Regression	When dependent and independent variables have an almost linear relationship, it generates an optimal result	Sensitive for the outliers and numeric value limitation
Support Vector Machine	The nonlinear problem of classification is solved with low power of computation and suitable for analyzing real-time data	Input data transformation results in difficulties while processing the data

collection, data preprocessing, analyzing of data, data training with different classifiers, respectively, and later testing the data. During the stage of preprocessing, the entire data are transformed into a useable format. The hybrid undersampling (negative class) and oversampling (positive class) techniques were performed using two different data distribution sets. In the stage of training, the classifier algorithm is fed with preprocessed data. Later, the testing data are evaluated to find the accuracy for detecting out the fraud related to a credit card. Finally, all the different models are evaluated based upon accuracy and their best performance. The legal ratio with a total number of fraud transactions is a subset and used to conclude which model performs better when tested in the real-time scenario.

3.1. Dataset. The source of the dataset is the UCI Machine Learning Repository. The dataset holds the information-related transaction conducted through credit cards as a default payment gateway of the different customers in Taiwan. The accuracy is probably compared to six different data mining techniques. The dataset has the detail of transaction which has occurred in the year 2015 and consists of 30000 different customer data and nearly 3 lakhs of transaction data. The characteristic of the dataset is multivariate, and its entire attributes are accurate and integer. The dataset seems to be highly unbalanced and more biased about positive class. It contains the continuous variable (numerical) as input variable was Principal Component Analysis. Altogether 30 different input features are used for training and testing the model. The detailed information related to the transaction's background and its features is not provided due to the issue of confidentiality. The preprocessing of the dataset is carried out using hybrid oversampling and undersampling techniques to achieve the two different sets of distribution in an unbalanced dataset.

The experimental setup used for performing fraud detection in credit cards is Python v3 language setup with i5 8th Gen Processor and 240 GB of SSD with 8 GB of DDR4 RAM with the processor variant of 1050 H which has the clock speed of 2.6 GHz–5.0 GHz with the turbo boost, and the frequency of RAM is 2565 MHz for training and testing the model in minimum duration of time.

3.2. Sequential Model. The sequential model generates its sequential value by estimating the input values for the series which can be time-series data. A 2D convolutional neural network is applied for passing 2D signals using more cost, time, and resources for gaining the state-of-art level of performance. It is easier to train the dataset through a sequential model as it requires minimum computation complexity and generates a better result.

3.3. Naive Bayes Classifier. Naive Bayes is the statistical method that relies on Bayesian theory, where the result is obtained based on the highest probability. It estimates the probability of the unknown value based upon the known value. The logic and prior knowledge can be applied to predict unknown probability. Naive Bayes mainly depends on binary classes and conditional probabilities.

$$\text{prob}\left(\text{class}_j|\text{feature}_k\right) = \frac{\text{prob}\left(\text{feature}_k|\text{class}_j\right) * \text{prob}\left(\text{class}_j\right)}{\text{prob}\left(\text{feature}_k\right)}, \tag{1}$$

$$\text{prob}\left(\text{feature}_k|\text{class}_j\right) = \prod_{j=1}^{m} \text{prob}\left(\text{feature}_k|\text{class}_j\right). \tag{2}$$

In equations (1) and (2), n indicates the maximum amount of features, $\text{prob}\left(\text{feature}_k|\text{class}_j\right)$ indicates the probability of generating feature value feature_k provided in class_j, and $\text{prob}\left(\text{feature}_k\right)$ and $\text{prob}\left(\text{class}_j\right)$ indicate the probability of occurrence of feature value feature_k and the occurrence of class class_j, respectively. This classifier was utilized for binary classification with the aid of the Bayesian principle.

3.4. K-Nearest Neighbor (KNN). The KNN classifier is an instance approach of learning where classification is conducted based on the measure of similarity calculated by

Table 2: Comparison of performance metrics among the utilized models.

Model	Accuracy (%)	Precision (%)	Recall (%)	$F1$-score (%)
Naive Bayes	96.1	92.4	91.86	92.13
Logistic Regression	94.8	93.16	93.07	93.11
K-Nearest Neighbor	95.89	93.78	91.42	92.58
Random Forest	97.58	96.5	96.7	96.60
Sequential CNN	92.3	90.3	90.43	90.36

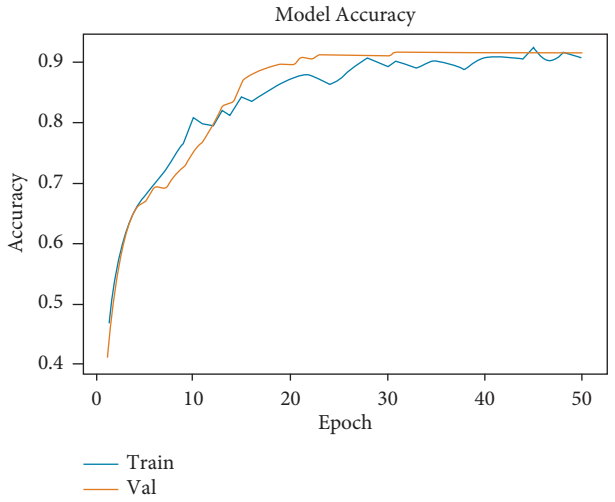

Figure 1: Representation of training accuracy vs. validation accuracy.

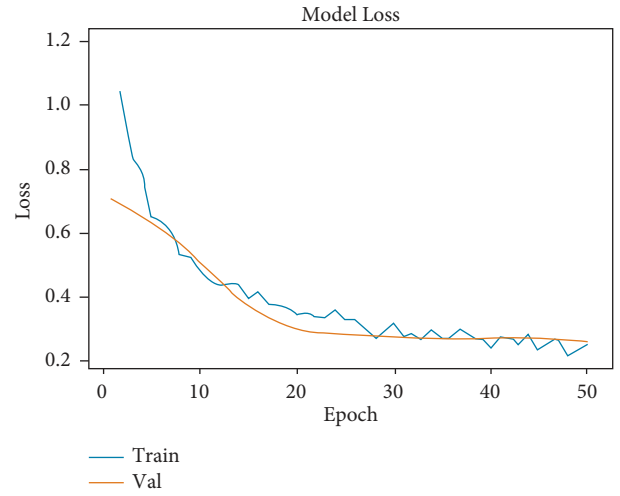

Figure 2: Representation of training loss vs. validation loss.

Manhattan or Euclidean and the Minkowski distance function. Manhattan or Euclidean function mainly deals with continuous variable, while the Minkowski deals with categorical data. The Euclidean function is used for measuring the distance in the KNN classifier. The Euclidean function (D_{ij}) between two vectors (X_i and X_j) is calculated by

$$\text{Dist}_{ij} = \sqrt{\sum_{l=1}^{m} \left(X_{il} - X_{jl}\right)^2}. \tag{3}$$

3.5. Logistic Regression. Logistic Regression is a functional approach for measuring the probability for binary classes based on particular or more features. It generates the best parameter for the sigmoid nonlinear function. The input vector (x) and the sigmoid function () are shown below.

The input data are a vector (z), and w is the best coefficient, when multiplied together and summarized to generate the targeted class classification classifier. If its value crosses 0.5, then it is known as 1, otherwise it is considered as 0. Then, the gradient ascent optimizer is applied in training for knowing the best performance of the classifier.

$$\text{Sig}_f(x) = \frac{1}{\left(1 + e^{-x}\right)},$$
$$x = w_0 z_0 + w_1 z_1 + w_2 z_2 + \cdots + w_n z_n. \tag{4}$$

3.6. Result. In this research, the sequential model and the other four models of a classifier based on KNN, Naive Bayes, Logistic Regression, and Support Vector Machine are developed. For evaluating all these classifier models, training is conducted using 70% of the entire dataset, while for testing and validating, 30% of the dataset is used. Accuracy, specificity, sensitivity, precision, and the Matthews correlation coefficient (MCC) with the rate of balance classification are applied for measuring the performance of all these classifier models. The performance of all these classifier models is evaluated. The sequential model visualizes the better performance. The technique of the sequential model generates superior performance for the evaluation metrics applied. It generates the highest value for precision and specificity. The obtained performance metrics are presented in Table 2.

The obtained results were plotted to visualize the comparison in terms of performance metrics. First, training accuracy vs. validation accuracy is represented in Figure 1. Second, training loss vs. validation loss is represented in Figure 2. Lastly, all performance metrics' comparison graph is represented in Figure 3.

4. Future Scope and Conclusion

The proposed methodology provides the information that Random Forest performs better than Sequential CNN. The drawback of this methodology is that anyone would expect Sequential CNN can outperform any of the conventional ML methodologies, but it is not happening here. It may happen because the dataset is not enough to train and identify the

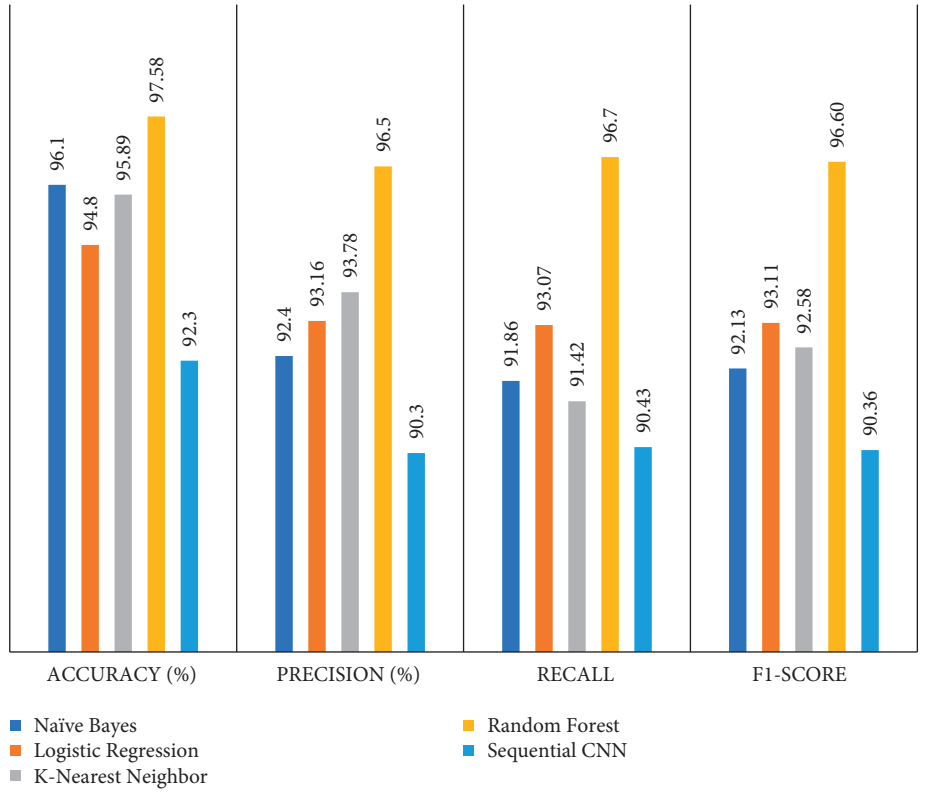

Figure 3: Comparison of performance metrics among the utilized models.

hidden patterns to predict the future or upcoming data and the initialization of weights was very random that might affect the training process. It can be further improved in two ways. The first way is to tune the hyperparameters through optimization, and the second method is to apply the transfer learning methodology so that the performance of the proposed methodology is improved to detect the fraud transaction through credit cards in the healthcare sector.

The study on fraud detection related to a credit card using deep learning and the machine learning techniques has been introduced in this paper. The different standard models such as Sequential Model, Decision Tree, Random Forest, and Naive Bayes are introduced and cast for empirical evaluation. The dataset related to a credit card is available publicly. Different standard models are trained and tested to generate the accuracy, and the model which performs better with stored and real-time data is identified. Sequential model and machine learning classifiers are trained and tested on the dataset, and their performance is evaluated with many relevant metrics for detecting fraud in credit cards. Our study indicates that online and offline transactions have different qualities when compared with the sequential pattern of earlier predicted fraud detection data.

The different algorithms presented in this paper can be extended towards the online learning approach of machine learning in the future. They can be investigated in both offline (collected data) and real-time scenario for obtaining better results with reasonable accuracy. The model of online learning will detect fraud cases in real time with the minimum time required for processing. This helps to predict the fraudulent transaction in an earlier stage (before conducted), which has positive impacts towards reducing the number of loss cases in the financial sector.

Data Availability

The data can be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully thank I-Cheng Yeh from the Department of Information Management of Chung Hua University, Taiwan, and the Department of Civil Engineering, Tamkang University, Taiwan, for the credit card fraud dataset and description related to it in the public domain.

References

[1] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: a comparative analysis," in *Proceedings of the 2017 International Conference on Computing Networking and Informatics (ICCNI)*, pp. 1–9, IEEE, Lagos, Nigeria, Oct. 2017.

[2] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: realistic modeling and a novel learning strategy," *IEEE transactions on neural*

networks and learning systems, vol. 29, no. 8, pp. 3784–3797, 2017.

[3] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in *Proceedings of the 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 1–6, IEEE, Zhuhai, China, March 2018.

[4] J. Jurgovsky, M. Granitzer, K. Ziegler et al., "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018.

[5] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit card fraud detection-machine learning methods," in *Proceeding of the 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–5, IEEE, East Sarajevo, Bosnia and Herzegovina, March 2019.

[6] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences*, vol. 557, pp. 317–331, 2021.

[7] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE access*, vol. 6, Article ID 14277, 2018.

[8] A. G. C. de Sá, A. C. M. Pereira, and G. L. Pappa, "A customized classification algorithm for credit card fraud detection," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 21–29, 2018.

[9] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. R. Rao, "Credit card fraud detection using machine learning," in *Proceeding of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1264–1270, IEEE, Madurai, India, May 2020.

[10] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit card fraud detection using pipeling and ensemble learning," *Procedia Computer Science*, vol. 173, pp. 104–112, 2020.

[11] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, 2013.

[12] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *International Journal of System Assurance Engineering and Management*, vol. 8, no. S2, pp. 937–953, 2017.

[13] A. Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection using hidden Markov model," *IEEE Transactions on Dependable and Secure Computing*, vol. 5, no. 1, pp. 37–48, 2008.

[14] M. V. Madhavan, A. Khamparia, D. Gupta, S. Pande, P. Tiwari, and M. S. Hossain, "Res-CovNet: an internet of medical health things driven COVID-19 framework using transfer learning," *Neural Computing & Applications*, vol. 33, pp. 1–14, 2021.

[15] C. Drummond and R. C. Holte, "C4.5, class imbalance, and cost sensitivity: why under-sampling beats oversampling," *Proc of the Icml Workshop on Learning from Imbalanced Datasets II*, vol. 45, pp. 1–8, 2003.

[16] M. V. Madhavan, S. Pande, P. Umekar, T. Mahore, and D. Kalyankar, "Comparative analysis of detection of email spam with the aid of machine learning approaches," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, Article ID 012113, 2021.

[17] J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1721–1732, 2008.

[18] A. Kundu, S. Panigrahi, S. Sural, and A. K. Majumdar, "BLAST-SSAHA hybridization for credit card fraud detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 6, no. 4, pp. 309–315, 2009.

[19] E. Shi, Y. Niu, M. Jakobsson, and R. Chow, "Implicit authentication through learning user behavior," *International Conference on Information Security*, vol. 6531, pp. 99–113, 2010.

[20] M. V. Madhavan, D. N. H. Thanh, A. Khamparia, S. Pande, R. Malik, and D. Gupta, "Recognition and classification of pomegranate leaves diseases by image processing and machine learning techniques," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 2939–2955, 2021.

[21] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Systems with Applications*, vol. 38, no. 10, Article ID 13057, 2011.

[22] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: a comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.

[23] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," in *Proceedings of the 2007 International Conference on Service Systems and Service Management*, pp. 1–4, IEEE, Chengdu, China, June 2007.

[24] Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and Logistic Regression," in *Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications (ISTA)*, pp. 315–319, IEEE, Istanbul, Turkey, June 2011.

[25] K. Chaudhary and B. Mallick, "Credit Card Fraud: the study of its impact and detection techniques," *International Journal of Computer Science and Network (IJCSN)*, vol. 1, no. 4, pp. 31–35, 2012.

[26] T. P. Bhatla, V. Prabhu, and A. Dua, "Understanding credit card frauds," *Cards Business Review# 2003-1*, Tata Consultancy Services, Mumbai, India, 2003.

[27] S. Stolfo, D. W. Fan, W. Lee, A. Prodromidis, and P. Chan, "Credit card fraud detection using meta-learning: issues and initial results," in *Proceeding of the AAAI-97 Workshop on Fraud Detection and Risk Management*, Melno Park, CA, USA, July 1997.

[28] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost-sensitive credit card fraud detection using Bayes minimum risk,"vol. 1, pp. 333–338, in *Proceeding of the 2013 12th International Conference onMachine Learning and Applications (ICMLA)*, vol. 1, , IEEE, Miami, FL, USA, Dec. 2013.

[29] J. K. F. Pun, "Improving credit card fraud detection using a meta-learning strategy,", Doctoral dissertation, University of Toronto , 2011.

[30] S. Sanober, I. Alam, S. Pande et al., "An enhanced secure deep learning algorithm for fraud detection in wireless communication," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6079582, 14 pages, 2021.