Credit Card Default Prediction: A Comparative Analysis of Machine Learning Models and Ensemble Techniques

Ajaypradeep Natarajsivam[1], K. Hemasree[2], M. Divija[3], D. Celeena Priyanka[4] and A. Gowthami[5]
{ajyapradeepn@mits.ac.in[1], hemasreehema79@gmail.com[2], mdivija2004@gmail.com[3], imcelina982@gmail.com[4], gowthamiayyannagari27@gmail.com[5]}

Assistant Professor, Department of CSE, Madanapalle Institute of Technology & Science, Andhra Pradesh, India[1]
Department of CSE, Madanapalle Institute of Technology & Science, Andhra Pradesh, India[2, 3, 4]

Abstract. CCDP is crucial for financial institutions to mitigate risks. While previous studies have primarily explored DT and AdaBoost models, limited research has assessed ensemble learning and DL techniques in this domain. Existing work often lacks transparency in feature selection, class imbalance handling, and computational efficiency analysis. This study evaluates multiple machine learning models, including LR, SVM, DT, RF, KNN, NB, ADB, ANN, and XGBoost. Unlike prior research, we employ outlier removal, feature scaling, and SMOTE to enhance model fairness. Our results show that Random Forest (85.97%) and XGBoost (85.06%) outperform AdaBoost (82%), contradicting previous findings. Additionally, execution time analysis highlights Random Forest and XGBoost as optimal trade-offs between accuracy and efficiency. This research provides a comprehensive evaluation of predictive models, offering valuable insights for improving credit risk assessment in banking.

Keywords: Credit Card Default Prediction, Machine Learning, Ensemble Learning, Random Forest, XGBoost, AdaBoost, ANN, SVM, SMOTE, Financial Risk Management.

1 Introduction

CCDP is relevant for financial institutions as it materially influences risk measurement and credit risk management. The spike in the use of credit has placed banks and lenders in a more precarious financial position as the threat of defaults rises with higher credit card use. However, conventional credit scoring algorithms are based on statistical approaches such as linear discriminant analysis and Logistic Regression (LR). However, complex, high-dimensional financial data often encounters difficult problems in applying these algorithms. The ability of ML models to uncover finer patterns and improve forecast quality has led to given popularity. This paper explores ample of machine learning (ML) approaches, such as ensembles learning and deep learning (DL), to improve the accuracy and efficiency of credit default prediction.

Most of the literature reviews related to credit risk modelling solutions focus on traditional ML methods like DT and AdaBoost. Although these models have achieved good performance, they are susceptible to over-fitting and are not robust for dealing with the imbalanced

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357982

datasets. In addition, DL methods, such as ANN, have been investigated in finance modeling but are not widely used because they are computationally intensive. Although ANNs are very powerful in learning complex relationships, yet they demand complex hyperparameter tuning and a vast dataset for best accuracy. We attempt to fill in this gap by investigating a variety of ML models and preprocessing them with feature scaling, outlier removing, and SMOTE to correct for class imbalance in the context of this work.

In CCDP, one of the most important issues is imbalanced dataset where the defaulters exist minority class. For traditional classifier, the bias of decision tends to be the majority-class in antifeed: biased classifiers. To overcome this, we use SMOTE that makes learning process better by creating synthetic minority samples. Furthermore, feature selection methods, such as correlation analysis and feature selection, are employed to eliminate irrelevant or insignificant features. Preprocessing methods are essential for the generalization of ML models to new data, which is important for the prediction performances.

Additionally, time analysis is carried out to assess the computational efficiency of the algorithms developed. Precision and recall In Table , the precision and recall numbers show that Random Forest and XGB outstrip other models with higher accuracy and computationally tractable. Unlike previous studies that preferred AdaBoost, our results reveal that in the context of credit defaults ensemble models exhibit significantly superior prediction accuracy.

The main aim of this study is to juxtapose machine learning approaches developed using ML algorithms for credit scoring application. We report such results to show how, by including the better of the advanced preprocessing methods and the tested classification models, we offer the finance community a more reliable and robust system for predicting defaults. This research demonstrates the significant trade-off between accuracy and computational efficiency that must be achieved for externally valid approaches in real banking systems. The findings of this study may help financial institutions to improve the quality of data used in their decision-making process in order to reduce the level of financial losses caused by credit card defaults.

2 Related Work

CCDP has been well-explored in a variety of ML methods from classical statistical to DL methods. Sun and Vasarhelyi [1] investigated DNN for credit card delinquency prediction.
Some alternative ML methods like gradient boosting, hybrid models, etc., have also been investigated in predicting credit default. Sayjadah et al. [2] applied ML techniques for default prediction, highlighting the importance of feature selection and algorithm choice. Teng and Lee [3] examined five ML algorithms and evaluated the significance of feature engineering in achieving better quality prediction. Syam and Sharma [4] discussed the broader impact of ML and AI in business applications, particularly in sales research, reinforcing their value in credit risk modeling. Gui [6] performed an experiment at the University of California, Los Angeles, and applied several algorithms to predict default on credit cards and demonstrated Random Forest is efficient for large data like our data. Alam et al. [7] explored the difficulties in learning from imbalanced corpus, and proposed resampling methods (e.g., SMOTE) to treat the model fairly, which we employ in this paper. Shabbir et al. [8] also investigated imbalanced credit datasets, comparing multiple models and emphasizing resampling effectiveness.

Several studies have also explored ANN and DL-based techniques. Ebiaredoh-Mienye et al. [5] proposed a stacked sparse autoencoder approach to improve ANN's predictive performance, showing its ability to capture complex relationships in financial data. However, DL models often require extensive tuning and large datasets for optimal performance, which can be a limitation in real-world applications. Chen and Zhang [12] combined K-means SMOTE with a backpropagation (BP) neural network, demonstrating improvements in CCDP by addressing class imbalance. Similarly, Gao et al. [13] introduced an XGBoost-LSTM model to leverage both feature importance and temporal dependencies in financial data, significantly enhancing predictive performance. Lundervold and Lundervold [22] provided a comprehensive overview of DL in other domains such as medical imaging, showing the transferability of DL approaches across disciplines. Choi et al. [23] compared optimizers for deep learning, insights from which are relevant for tuning DL models in CCDP.

In addition to supervised learning models, hybrid approaches and real-time fraud detection techniques have been explored. Tanouz et al. [9] applied ML methods for credit card fraud detection at ICICCS 2021, demonstrating their potential in handling high-volume transactional data. Varmedja et al. [14] examined ML-based methods for fraud detection, emphasizing challenges in practical deployment. Dornadula and Geetha [15] examined multiple ML models for fraud detection, emphasizing the advantages of ensemble learning methods. Thennakoon et al. [16] implemented real-time fraud detection systems using ML, demonstrating the practicality of AI-driven financial risk management. Maniraj et al. [17] highlighted how combining ML with data science pipelines improves fraud detection accuracy, while Yee et al. [18] approached fraud detection as a data mining problem. Sailusha et al. [19] proposed ML-based systems for fraud detection, validating their effectiveness on real-world datasets. Lakshmi and Kavilla [20] developed ML systems specifically tailored for fraud detection in credit card usage, while Trivedi et al. [21] proposed an efficient fraud detection model that optimizes computational costs. Adepoju et al. [10] conducted a comparative evaluation of fraud detection models, reinforcing the effectiveness of ensemble learning techniques in financial applications. Alfaiz and Fati [11] further enhanced fraud detection models using ML techniques, reporting improved classification accuracy. Despite these advancements, existing studies often lack a direct comparison of multiple classification algorithms with extensive preprocessing techniques.

Our research builds on previous work by implementing a comprehensive evaluation of multiple models. Additionally, we integrate outlier removal, feature scaling, and SMOTE to ensure a fair assessment of all models. By addressing these limitations, our study provides a more balanced comparison of ML models for CCDP, identifying the best trade-offs between accuracy and computational efficiency.

3 Methods and Materials

3.1 Dataset Description

This study utilizes the UCI CC Dataset, which consists of 30,000 customer records and 24 attributes. The dataset contains financial variables such as credit limit, past payment history, bill amounts, and delinquency records. The target variable, "default. payment. next. month," is a binary indicator in which 0 denotes no default and 1 denotes a default. The dataset was selected due to its extensive use in credit risk modeling and its availability in the UCI ML Repository.

3.2 Data Preprocessing

Improving model performance and guaranteeing significant predictions depend on proper data preparation. The actions listed below were taken.:

- **Handling Missing Values:** No missing values were found in the dataset, so imputation was not required.
- **Feature Selection:** Correlation analysis was conducted to remove redundant features that could introduce noise into the model.

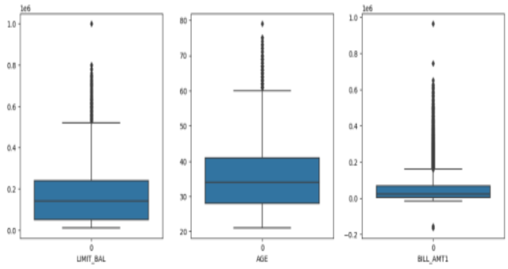

Fig. 1. Outlier Detection for Limit balance, age and bill amount.

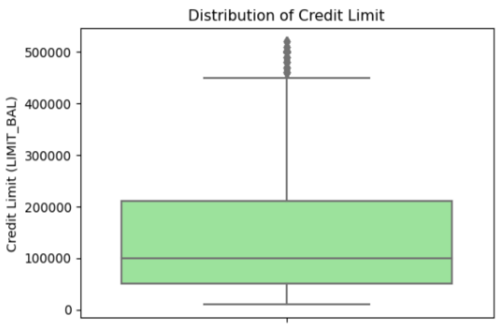

Fig. 2. Outlier detection on distribution of credit limit.

Fig 1 shows the Outlier Detection for Limit balance, age and bill amount and Fig. 2. Shows the Outlier detection on distribution of credit limit.

- **Standard Scaler** was applied to numerical features, preventing bias in distance-based models like KNN and SVM.
- **Class Imbalance Handling:** The dataset exhibited a class imbalance, with significantly more non- defaulters than defaulters. To address this, the SMOTE was used to generate synthetic data points, ensuring that models do not favor the majority class.

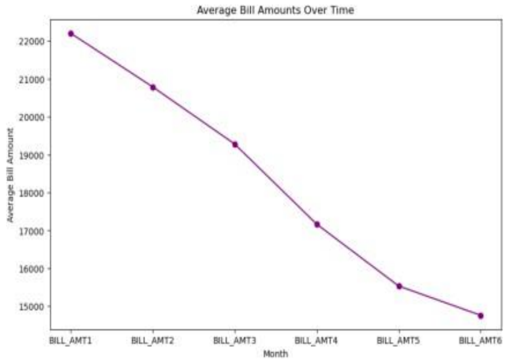

Fig. 3. Average bill amount over time.

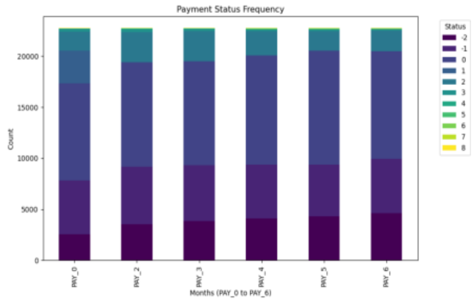

Fig. 4. Payment status frequency.

Fig 3 shows the Average bill amount over time and Fig 4 shows the Payment status frequency.

3.3 ML Models

To evaluate CCDP, we implemented the following ML models:

- LR – A statistical model that estimates the probability of default based on input features.
- DT – A tree-based model that classifies customers based on their financial attributes using hierarchical decision rules.
- RF – An ensemble method that constructs multiple decision trees to enhance prediction robustness and reduce overfitting.
- SVM – A margin-based classifier that finds an optimal hyperplane for binary classification.
- KNN – A distance-based model that classifies instances based on their similarity to neighboring data points.
- NB – A probabilistic approach that applies Bayes' theorem to predict class probabilities.
- AdaBoost – A boosting method that sequentially improves weak classifiers by focusing on misclassified instances.

- ANN – A multi-layered DL model capable of capturing complex nonlinear patterns.
- XGBoost – A powerful boosting algorithm known for its superior handling of structured tabular data.

Each model was trained and optimized using hyperparameter tuning to maximize predictive accuracy while avoiding overfitting.

3.4 Model Training and Evaluation

To guarantee generalisation, we divided the dataset into 90% training and 10% testing set, preserving the class ratio. Performance: Model performance was assessed using Performance metrics.

Execution time: Measures the effectiveness of each model for computational time, which is important in real-time analysis in financial enterprises.

Experimental results showed that Random Forest (85.97%) and XGBoost (85.06%) achieved the best accuracy compared to 82% of AdaBoost, the previous state-of-the-art model in related works. Moreover, execution time comparison revealed that SVM system took much more time to calculate the results compared to the ensemble models, so that are more suitable trade-off between accuracy and time.

4 Results and Analysis

4.1 Model Performance Evaluation

The performances of Random Forest (85.97%) and XGBoost (85.06%) are the most excellent with a great deal of pretend above that of traditional models such as Logistic Regression (69.24%) and DT (69.24%). The accuracy comparison details can be seen below:

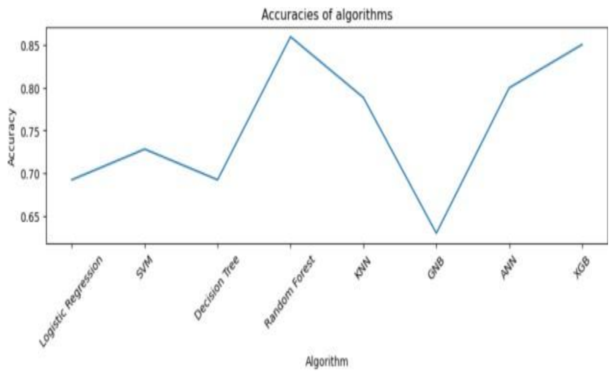

Fig. 5. Accuracies of algorithms.

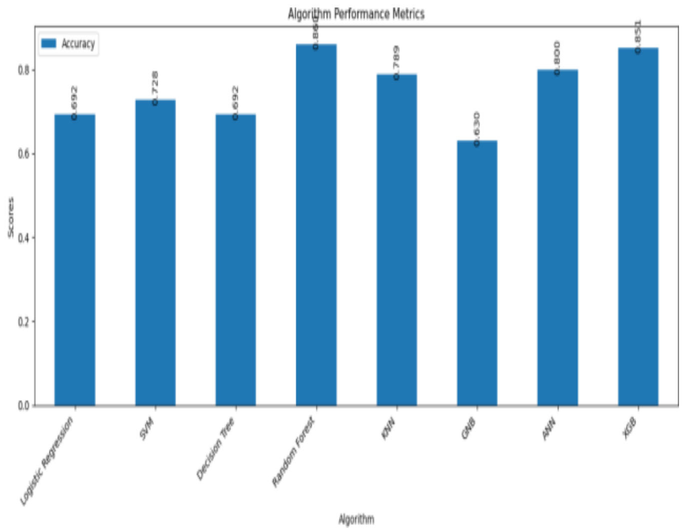

Fig. 6. Algorithms performance metrics.

The confusion matrices further highlight how each model performed in correctly classifying defaulters and non-defaulters. Fig 5 shows the Accuracies of algorithms and fig 6 shows the Algorithms performance metrics.

4.2 Confusion Matrix Analysis

The confusion matrices provide deeper insights into model performance. The Random Forest model exhibited a balanced classification, reducing false positives and false negatives significantly:

Random Forest:

TP (1512), TN (1509)

FP (245), FN (248)

XGBoost:

TP (1422), TN (1567)

FP (335), FN (190)

SVM and Logistic:

Regression had higher false negatives, indicating they struggled with identifying actual defaulters. In table 1, Naïve Bayes showed the poorest classification, likely due to the assumption of feature independence, which is unrealistic for financial data.

Table 1. Accuracies of our models.

Algorithm	Acc (%)
Lr	69.24
SVM	72.82
DT	69.24
RF	85.97
KNN	78.86
NB	62.97
ANN	80.00
XGB	85.06

4.3 Execution Time Analysis

Execution time is a crucial factor for real-world applications, especially in financial institutions where rapid decision-making is required. The computational time of each model is shown below in table 2:

Table 2. Execution time in seconds.

Algorithm	Execution Time (Seconds)
LO R	1.08
SVM	115.19
DT	3.16
RF	62.23
KNN	0.02
NB	0.04
XGB	1.29

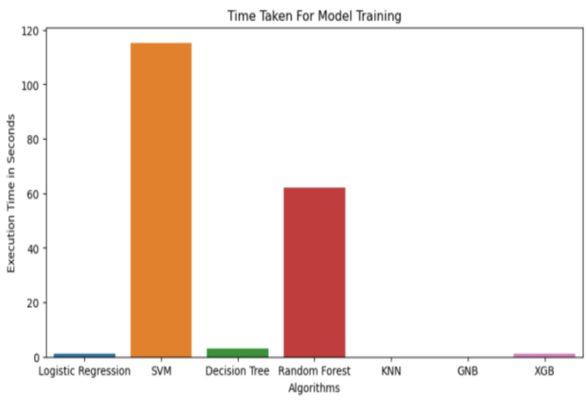

Fig. 7. Time taken for model training.

In fig 7, SVM had the highest execution time (115.19 seconds), making it impractical for large-scale deployment. KNN was the fastest (0.02s) but had lower predictive power, making it unsuitable for real-world use.

4.4 Comparative Performance Analysis

RF and XGB emerged as the best models due to their high accuracy, low false negatives, and acceptable execution times. ANN performed well (80%) but required more computational resources. SVM, despite good accuracy, was computationally expensive. Logistic Regression and Decision Tree had moderate accuracy but lacked robustness in handling class imbalance. Naïve Bayes performed poorly due to the complex interdependencies among financial variables.

4.5 Key Findings and Insights

Class imbalance significantly impacted model performance; SMOTE effectively mitigated this issue. Computational efficiency is a critical consideration for real-time financial applications. Future improvements may involve hyperparameter tuning, feature selection optimizations, and DL techniques such as CNNs or transformers for better accuracy.

5 Conclusion

This study explored various ML models for CCDP, comparing traditional classifiers, ensemble learning techniques, and DL models. Through comprehensive data preprocessing, including outlier removal, feature scaling, and SMOTE for class imbalance handling, we ensured fair model evaluation. Our results demonstrated that Random Forest (85.97%) and XGBoost (85.06%) outperformed other models, proving their effectiveness in financial risk assessment. While ANN provided competitive accuracy (80%), its computational complexity makes it less practical for large-scale real-time applications. On the other hand, SVM, despite being a strong classifier, had the highest execution time (115.19s), limiting its applicability in time-sensitive environments.

The results reveal that the ensemble learning methods present the optimal trade-off between accuracy and speed to predict credit card defaults. This work could be expanded by applying more hyperparameter tuning, more oversampling techniques and different DL architectures like transformers and the RNNs. It may also be advantageous to include real time financial transaction data into it.

6 Discussion and Future Work

6.1 Discussion

The findings of this study reveal that ensemble learning methods, especially Random Forest (85.97%) and XGBoost (85.06%) eclipse traditional ML models for the default prediction of credit cards. They are able to capture strong trends in financial time series and by combining multiple weak learners can reduce over-fitting. On the other hand, logistic regression (69.24%) and decision tree (69.24%) presented low accuracy, indicating that simple models have difficulties coping with high-dimensional financial data. ANN produces 80% accuracy, demonstrating its capacity to grasp complex relations, however the training complexity and computational cost make it hard to scale.

Execution time analysis revealed that SVM required the longest computational time (115.19s), making it impractical for real-time deployment. Conversely, KNN was the fastest (0.02s) but lacked predictive power, proving unsuitable for financial risk modeling. These findings emphasize the importance of selecting a balanced trade-off between accuracy and efficiency when deploying ML models in credit risk assessment. Additionally, class imbalance significantly influenced the performance of models, reinforcing the necessity of using SMOTE to enhance predictive fairness.

6.2 Future Work

Although this study provides valuable insights into CCDP using ML, several areas remain open for improvement. Hyperparameter tuning and feature selection techniques could further optimize model performance and reduce computational costs. Future research can explore the application of DL architectures such as LSTMs, transformers, and hybrid models (e.g., XGBoost-LSTM) to enhance predictive accuracy while maintaining interpretability.

In addition, real world financial applications must process data in real time, which calls for the use of streaming ML methods that can adapt to evolving credit risk patterns. There is also room for further development of more interpretable and interpretable models to help with regulation in the financial decision making. Lastly, researching this using multi-source financial data and macroeconomic associated factors, this could enhance prediction accuracy, which can even make ML based credit risk prediction systems more reliable and usable in banking and financial industry.

Acknowledgment

The authors express their sincere gratitude to the UCI ML Repository for providing the publicly available CCD dataset, which served as the foundation for this research. We also acknowledge the contributions of researchers whose prior work in credit risk assessment and ML applications provided valuable insights for this study.

Additionally, we extend our appreciation to our academic mentors, colleagues, and peers for their constructive feedback and guidance throughout the research process. Their insights and discussions significantly contributed to refining our methodologies and improving the quality of our analysis. Finally, we acknowledge the support of our institution and faculty members, whose encouragement and resources facilitated the successful completion of this study.

References

[1] T. Sun, M.A. Vasarhelyi, Predicting credit card delinquencies: an application of deep neural networks, Intell. Syst. Account. Finance Manag. 25 (4) (2018) 174–189.

[2] Y. Sayjadah, I.A.T. Hashem, F. Alotaibi, K.A. Kasmiran, Credit card default prediction using machine learning techniques, in: 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), IEEE, October 2018, pp. 1–4.

[3] H.W. Teng, M. Lee, Estimation procedures of using five alternative machine learning methods for predicting credit card default, Rev. Pac. Basin Financ. Mark. Policies 22 (3) (2019) 1950021.

[4] N. Syam, A. Sharma, Waiting for a sales renaissance in the fourth industrial revolution: machine learning and artificial intelligence in sales research and practice, Ind. Market. Manag.

69 (2018) 135–146.
[5] S.A. Ebiaredoh-Mienye, E. Esenogho, T.G. Swart, Artificial neural network technique for improving prediction of credit card default: a stacked sparse autoencoder approach, Int. J. Electr. Comput. Eng. 11 (5) (2021) 4392.
[6] L. Gui, Application of Machine Learning Algorithms in Predicting Credit Card Default Payment, University of California, Los Angeles, 2019. Master's thesis.
[7] T.M. Alam, K. Shaukat, I.A. Hameed, S. Luo, M.U. Sarwar,
[8] S. Shabbir, J. Li, M. Khushi, An investigation of credit card default prediction in the imbalanced datasets, IEEE Access 8 (2020) 201173–201198.
[9] Tanouz, R.R. Subramanian, D. Eswar, G.P. Reddy, A.R. Kumar, C.V. Praneeth, Credit card fraud detection using machine learning, in: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, May 2021, pp. 967–972.
[10] O. Adepoju, J. Wosowei, H. Jaiman, Comparative evaluation of credit card fraud detection using machine learning techniques, in: 2019 Global Conference for Advancement in Technology (GCAT), IEEE, 2019, pp. 1–6.
[11] N.S. Alfaiz, S.M. Fati, Enhanced credit card fraud detection model using machine learning, Electronics 11 (4) (2022) 662.
[12] Y. Chen, R. Zhang, Research on credit card default prediction based on k-means SMOTE and BP neural network, Complexity (2021) 1–13.
[13] J. Gao, W. Sun, X. Sui, Research on default prediction for credit card users based on XGBoost-LSTM model, Discrete Dynam Nat. Soc. (2021) 1–13.
[14] Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, A. Anderla, Credit card fraud detection-machine learning methods, in: 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE, March 2019, pp. 1–5.
[15] V.N. Dornadula, S. Geetha, Credit card fraud detection using machine learning algorithms, Procedia Comput. Sci. 165 (2019) 631–641.
[16] Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, N. Kuruwitaarachchi, Real-time credit card fraud detection using machine learning, in: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, January 2019, pp. 488–493.
[17] S.P. Maniraj, A. Saini, S. Ahmed, S. Sarkar, Credit card fraud detection using machine learning and data science, Int. J. Eng. Res. 8 (9) (2019) 110–115.
[18] O.S. Yee, S. Sagadevan, N.H.A.H. Malim, Credit card fraud detection using machine learning as data mining technique, J. Telecommun. Electron. Comput. Eng. 10 (1–4) (2018) 23–27.
[19] R. Sailusha, V. Gnaneswar, R. Ramesh, G.R. Rao, Credit card fraud detection using machine learning, in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, May 2020, pp. 1264–1270.
[20] S.V.S.S. Lakshmi, S.D. Kavilla, Machine learning for credit card fraud detection system, Int. J. Appl. Eng. Res. 13 (24) (2018) 16819–16824.
[21] N.K. Trivedi, S. Simaiya, U.K. Lilhore, S.K. Sharma, an efficient credit card fraud detection model based on machine learning methods, International Journal of Advanced Science and Technology 29 (5) (2020) 3414–3424.
[22] A.S. Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, Z. Med. Phys. 29(2) (2019) 102–127.
[23] Choi, C.J. Shallue, Z. Nado, J. Lee, C.J. Maddison, G.E. Dahl, On empirical comparisons of optimizers for deep learning. arXiv Preprint arXiv:1910.05446, 2019.