International Journal of Computer and Communication Technology

Volume 8 | Issue 4 Article 4

November 2022

Credit Card Fraud Detection Using Logistic Regression and Synthetic Minority Oversampling Technique (SMOTE) Approach

Nrusingha Tripathy
Siksha 'O' Anusandhan, nrusinghatripathy654@gmail.com

Subrat Kumar Nayak
Siksha 'O'Anusandhan, subratsilicon28@gmail.com

Julius Femi Godslove
juliusgodslove88@gmail.com

Ibanga Kpereobong Friday
kpereib01@gmail.com

Sasanka Sekhar Dalai
Siksha 'O'Anusandhan, sasanka.sekhar.dalai@gmail.com

Credit Card Fraud Detection Using Logistic Regressionand Synthetic Minority Oversampling Technique (SMOTE) Approach

Nrusingha Tripathy[1*], Subrat Kumar Nayak[1], Julius Godslove Femi[2],Ibanga Kpereobong Friday[2],
Sasanka Sekhar Dalai[3]
Department of Computer Science & Engineering, Institute of Technical Education and Research,
Siksha 'O'Anusandhan (Deemed to be) University, Bhubaneswar, India-751030 [1*,1,2,2*,3]
nrusinghatripathy654@gmail.com[1*], subratsilicon28@gmail.com[1],
juliusgodslove88@gmail.com[2*],kpereib01@gmail.com[2], sasanka.sekhar.dalai@gmail.com[3]

***Abstract*-**Financial fraud is a serious threat that is expanding effects on the financial sector. The use of credit cards is growing as digitization and internet transactions advance daily. Many plastic cards in circulation throughout the world are like a gold mine. Credit card losses are predicted to cost financial service providers $40 billion globally by 2027, up from $27.85 billion in 2018.The emergence of electronic transactions is partially to blame for this increase in losses. Imagine that 1.5 billion credit cards are currently in use in the US alone, with the average American having more than three cards. While there are an amazing 22.11 billion plastic cards in use worldwide. Recognising counterfeit credit card transactions is difficult, as it prevents credit card firms' consumers from being charged for goods they did not buy. The most common issues in today's culture are credit card scams. This kind of fraud typically happens when someone uses someone else's credit card details. Credit card fraud detection uses transaction data attributes to identify credit card fraud, which can save significant financial losses and affluence the burden on the police. The detection of credit card fraud has three difficulties: uneven data, an abundance of unseen variables, and the selection of an appropriate threshold to improve the models' reliability. This study employs a modified Logistic Regression (LR) model to detect credit card fraud in order to get over the preceding difficulties. The dataset sampling strategy, variable choice, and detection methods employed all have a significant impact on the effectiveness of fraud detection in credit card transactions. This study investigates logistic regression on extremely biased credit card fraud data. The accuracy of the logistic regression technique will be closer to 0.98%; with this accuracy, frauds may be easily detected. The fact that LR receives the highest classifier score illustrates how well LR predicts credit card theft.

***Keywords*-**fraud detection; credit card fraud; logistic regression; Synthetic Minority Oversampling Technique (SMOTE)

I. INTRODUCTION

Fraud is defined as unlawful or illegal deceit meant to produce financial or personal benefit or to cause harm to another person without necessarily having legal repercussions. Systems for fraud prevention and fraud detection are the two primary tools for preventing frauds and losses brought on by fraudulent actions. Fraud prevention is a proactive method designed to stop fraud from happening. When criminals circumvent fraud protection measures and initiate a fraudulent transaction, fraud detection systems are activated [1,2]. Nobody is able to determine whether a fraudulent transaction has evaded the security measures. Because of this, the goal of systems that detect fraud is to identify fraudulent activities as soon as feasible after the fraudster has begun to carry out an illicit transaction by examining each transaction for the possibility of fraud despite the preventative measures. You

may read a review of the fraud detection systems in with advancements in communication networks and information technologies, fraud is spreading over the globe and causing significant financial losses. Online media like the Internet are the most popular, yet there are numerous other media that may be used to conduct fraud, including mail, wire, phone, and the Internet [3-7]. Fraudulent activity on the internet is increasing quickly as a result of the web's global accessibility and users' ease of concealing their identity and location during online transactions. Also, with the expansion of internetworking channels' capacity, fraudsters now have the opportunity to collaborate and exchange information globally to create fraud networks. Due to their nature, online credit card scams, for example, have become the most prevalent type of fraud.

Credit card fraud may be carried out in a number of methods, such as outright theft, application fraud, the use of fictitious cards, never receiving issue (NRI), and online fraud (when the cardholder is not present). Remote online fraud just needs the card details to be committed [8-11]. At the moment of purchase, neither a manual signature nor a PIN or card imprint are necessary. Online frauds (internet and mail order frauds) are continually growing in both quantity and number of transactions, despite the fact that protection measures like CHIP&PIN reduce fraudulent activity through simple theft, counterfeit cards, and NRI. With the use of credit cards becoming more and more widespread, there have been an increasing number of financial losses as a result of fraud using credit cards. Many publications detailed significant losses in numerous nations [12, 13].

Online scams accounted for roughly 50% of all credit card fraud losses in 2008, according to Visa data on European countries. Detecting credit card fraud is a very tough yet common problem to handle. First of all, the transaction being committed only contains a little amount of data, such as the transaction amount, date, and time, address, merchant category code (MCC), and merchant's acquirer number. It is quite challenging to match a pattern since there are millions of locations and online stores where a credit card might be used. Moreover, fraudsters may have engaged in previous transactions that appear to follow a pattern of typical (legal) conduct [14]. The problem also has a lot of restrictions. First of all, the characteristics of legitimate and fraudulent activity are always evolving. Second, the interchange of ideas in fraud detection, particularly in credit card fraud detection, is highly constrained owing to security and privacy considerations, which makes the development of novel fraud detection approaches more challenging. Lastly, it is challenging to evaluate the outcomes since data sets are frequently withheld and filtered. There is no prospect of benchmarking for the developed models as a result of this issue. Even some of the research use data that was artificially created [15-19]. Overall summary about the fraud detection is given,

- Every day, enormous amounts of data a re collected, and the model developmen t must be quick enough to detect scams in time.
- Data availability since the majority of t he transactions (99.8%) are private mak es it difficult to discover the fraudulent ones due to imbalanced data, which me ans that the majority of transactions are not fraudulent.
- Misclassified Data is another significant problem since not all fraudulent
- Transactions are discovered and report ed.

- The fraudsters utilised adaptive method s against the model.

II. LITERATURE SURVEY

We cannot disregard the losses associated with electronic commerce, even when services make electronic payments more convenient, frictionless, adequate, and easy to use. Businesses and banks offer reliable security alternatives for their use [20–25]. These issues lead to a shift in the misleading strategies utilised by con artists. Therefore, it is crucial to establish prevention and detection techniques[26, 27]. In order to successfully combat fraud, it is crucial to understand how it is carried out. The fraud strategy itself determines the instrument used to detect credit card fraud. Send the transaction information to the confirmation module, and it will classify the transactions as fraudulent or not. If it is shown to be fake, it will be rejected [28–33].

The transaction is approved in all other cases. Artificial intelligence and statistical data analysis are two fraud detection methods that may be utilised to differentiate between the two. Data mining is a technique used in AI that can categorise, organise, and segment data so that it may be searched through

millions of transactions to look for trends and identify fraud. A method for automatically identifying fraud features is machine learning. Using both prevention and detection to combat fraud is one strategy [34-39]. The main objectives of fraud detection and prevention are to distinguish between genuine and fraudulent transactions and to stop fraud. To assess if a transaction is fraudulent or not, the user's pattern and behaviour are examined using previous data [40, 41]. Fraud detection steps in when the system is unable to identify and stop fraudulent actions.depending on the features of deceitful and legal activities, are assessed as genuine or fraudulent, whereas outliers' transactions are recognised as potential fraudulent transactions in unsupervised fraud detection systems. There can be dialogue between supervised and unsupervised machine learning techniques. Numerous studies have been conducted on various approaches for tackling the card fraud identification issue. The DT, K-means Clustering, and ANN are some of these techniques[42].In this article, we'll talk about several kinds of credit card fraud and practical ways to spot it. highlighting difficulties that cardholders have, such as card issuers, fraudsters, latest information about them, credit cards, and offering some protective strategies to enable cardholders to engage in fraudulent activity. Data mining was used to compare the methods used in credit card theft by T. Abdul Razak and G. Najeeb Ahmed [5]. One of the crucial tasks is developing an accurate and user-friendly credit card risk monitoring system in order to successfully reduce the risk of store management levels. The purpose of this article is to uncover the user patterns that lead to the detection of high-level fraud instances.

III. METHODOLOGY

According to Fig 1 data is first loaded into the Jupyter notebook. In order to make the data easier to handle, we transform the data once it has been loaded into a data frame using pandas. We visualise the data after it has been loaded. We use data frame because we first need to know how our data is organised [43]. In addition, we need to know how the data is distributed, therefore we plot the data using head () to display the top five rows of the data

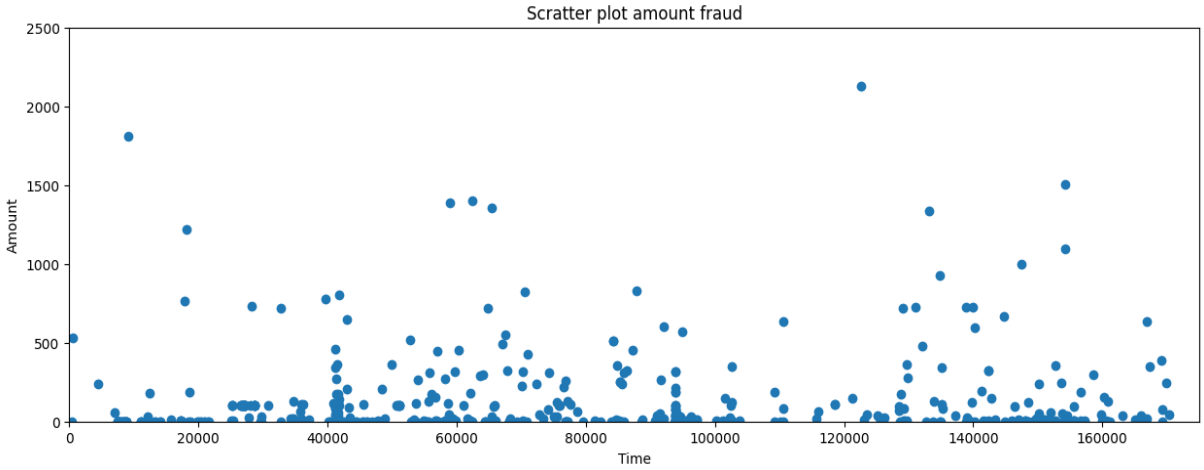

Fig 1. Scatter plot amount of fraud

Because of the data's extreme imbalance, the data must be under sampled [44]. Due to the extreme asymmetry in our data, we are under sampling it. The number of transactions that are not fraudulent is labelled as 0, whereas the number of fraudulent transactions is labelled as 1. There were 284315 legitimate transactions, whereas there were just 492 fraudulent ones. Under sampling is a much superior strategy to obtain the ideal and desired result since if we oversample our data, the inclusionof over 284000 dummy components would undoubtedly alter our results by a significant margin and it will be greatly prejudiced as non-fraudulent.Fig 2 shows the heat map of correlation of the futures. Fig 3 and Fig 4 describes the confusion matrix of logistic regression and Synthetic Minority Oversampling Technique (SMOTE)approach.

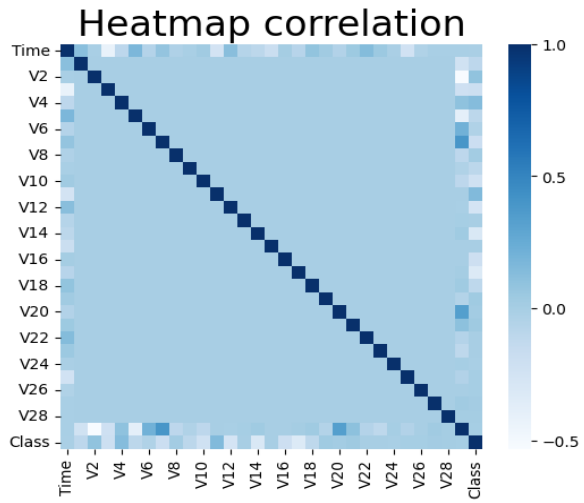

Fig 2.correlation of the futures

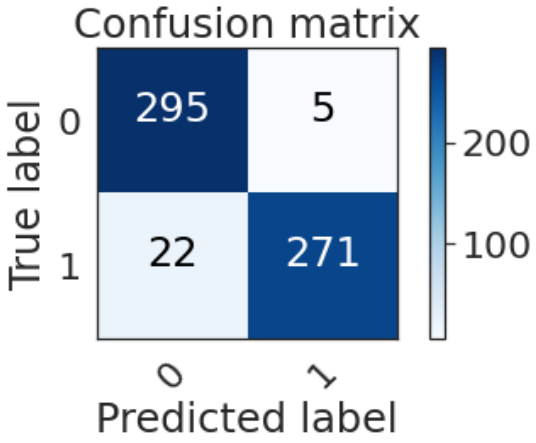

Fig 3. Confusion matrix of training model

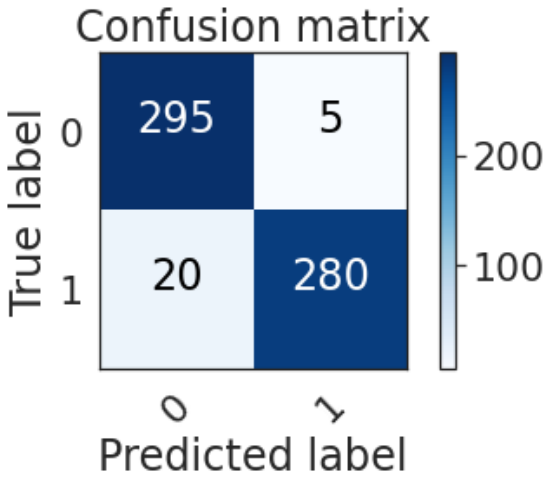

Fig 4. Confusion matrix after under sampling and SMOTE

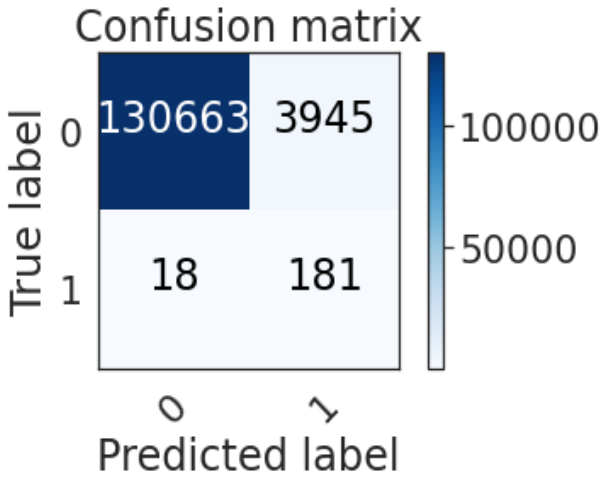

Fig 5. Confusion matrix after testing model

IV. RESULTANALYSIS

In general, it is more expensive to lose a fraudulent transaction than to wrongly label a legitimate transaction as fraudulent. Finding the right balance when training your model and moving forward appropriately is one of the problems.We frequently achieve accuracy of our training model of more than 95% using random data.Fig 5 shows the confusion matrix of our test data set. Table 1 shows the Mean Absolute Error(MAE), Mean Absolute Percentage Error(MAPE), Mean Square Error(MSE), R-Square values of Logistic Regression and Synthetic Minority Oversampling Technique (SMOTE) fortrain dataset, also we calculated the overall error score for test dataset.

Table 1. Error Calculation Chart

Name	Logistic Regression(Train data)	SMOTE (Train data)	Overall Score (Test data)
MAE	0.041	0.057	0.042
MAPE	0.035	0.0001	0.0001
MSE	0.041	0.057	0.042
R-Square	0.831	0.379	0.276

V. CONCLUSION

Credit card theft is causing a lot of anxiety worldwide. Fraud causes tremendous financial losses across the world. Credit card companies were compelled to spend money on developing and establishing strategies to detect and reduce fraud as a result. The major goal of this project is to provide appropriate technologies that credit card providers may employ to detect fraudulent transactions more quickly and inexpensively. Several machine learning methods are contrasted, including Artificial Neural Networks, Logical Regression, Decision Trees, Random Forest, Artificial Neural Networks, K-Nearest Neighbours, and K-means clustering. Since no two situations are identical, a situation-based approach may

be used to select the scenario that best fits a specific circumstance.

Every fraud detection method included in this article has benefits and drawbacks.

The researchers use a range of performance metrics (techniques) and algorithms to predict and identify fraudulent transactions. In order to determine the proper weight to assign to cost factors, testing accuracy, and fraud detection accuracy, studies are being relaunched and encouraged. Through such efforts, the researchers will be able to develop a hybrid approach that is most effective at identifying unauthorised credit card transactions.

Reference

[1] R.J. Bolton and D.J. Hand, "Statistical fraud detection: A review", Statistical Science, vol. 28, no. 3, pp. 235-255, 2002.

[2] Y. Kou, C.-T. Lu, S. Sirwongwattana and Y.-P. Huang, "Survey of fraud detection techniques", Proceedings of the 2004 IEEE International Conference on Networking Sensing and Control, 2004.

[3] C. Phua, V. Lee, K. Smith and R. Gayler, "A comprehensive survey of data mining-based fraud detection research", Artificial Intelligence Review, 2005.

[4] Y. Sahin and E. Duman, "An overview of business domains where fraud can take place and a survey of various fraud detection techniques", Proceedings of the 1st International Symposium on Computing in Science and Engineering, 2010.

[5] Razak, T.A. and Ahmed, G.N., 2014. A comparative analysis on credit card fraud techniques using data mining. International Journal of Data Mining Techniques and Applications, Integrated Intelligent Research (IIR), *3*(2), pp.398-400.

[6] R. Chen, M. Chiu, Y. Huang and L. Chen, "Detecting credit card fraud by using questionnaire-responded transaction model based on SVMs", Proceedings of IDEAL2004, 2004.

[7] D. J. Hand and G. Blunt, "Prospecting gems in credit card data", IMA Journal of Management Mathematics, vol. 12, 2001.

[8] N. Mahmoudi and E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis", Elsevier Expert System with Application, pp. 2510-2516, 2015.

[9] N. Halvaiee and M. Akbari, "A novel model for credit card fraud detection using Artificial Immune System", Elsevier Applied Soft Computing, pp. 40-49, 2014.

[10] M. Zareapoor, K. Seeja and M. Alam, Analysis of credit cardfraud detection techniques: based on certain design criteria, pp. 35-42, 2012.

[11] V. Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, et al., "APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network based Extensions", ELSEVIER Decision Support Systems, pp. 38-48, 2015.

[12] Y. Sachin and E. Duman, "Detecting Credit Card Fraud by Decision Tree and Support Vector Machine", Proceedings of the international multi-Conference of Engineers and Computer Scientists, pp. 1-6, 2011.

[13] J. West and M. Bhattacharya, "Intelligent Financial Fraud Detection: A Comprehensive Review", ELSEVIER Computer & Security, pp. 47-66, 2016.

[14] P. Ravisankar, V. Ravi, G. Rao and I. Bose, "Detection of financial statement fraud and feature selection using datamining techniques", Elsevier Decision Support Systems, pp. 491-500, 2011.

[15] M. Zareapoor and P. Shamsolmoali, "Application of Credit card Fraud Detection: Based on Bagging Ensemble Classifier", Elsevier International Conference on Intelligent Computing Communication & Convergence, pp. 679-685, 2015.

[16] G. Rushin, C. Stancil, M. Sun, S. Adams and P. Beling, "Horse Race Analysis in Credit card Fraud- Deep Learning Logistic Regresion and Gradient Boosted Tree", *IEEE*, pp. 117-121, 2017.

[17] A. Srivastava, A. Kundu, S. Sural and A. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model", IEEE Transaction on Dependable and Secure Computing, pp. 37-48, 2008.

[18] M. Zeager, A. Sridhar, N. Fogal, S. Adams, D. Brown and P. Beling, "Adversarial Learing in

Credit card Fraud Detection", IEEE, pp. 112-116, 2017.

[19] PATI, ABHILASH; Parhi, Manoranjan; and Pattanayak, Binod Kumar (2022) "IoT-Fog-Edge-Cloud Computing Simulation Tools, A Systematic Review," International Journal of Smart Sensor and Adhoc Network: Vol. 3 : Iss. 2 , Article 3.

[20] S. Kumar Nayak, A.Kumar Nayak, S. .Mishra, and P. Mohanty, “Deep Learning Approaches for Speech Command Recognition in a Low Resource KUI Language”, Int J Intell Syst Appl Eng, vol. 11, no. 2, pp. 377–386, Feb. 2023.

[21] Pati, Abhilash, Manoranjan Parhi, and Binod Kumar Pattanayak. "COVID-19 pandemic analysis and prediction using machine learning approaches in India." In Advances in Intelligent Computing and Communication: Proceedings of ICAC 2020, pp. 307-316. Springer Singapore, 2021.

[22] Y. Sahin and E. Duman, “Detecting credit card fraud by decision trees and support vector machines”, Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 2011.

[23] V. Bhusari & S. Patil. (2011). Application of hidden markov model in credit card fraud detection. International Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.6.

[24] A. Chittur. Model generation for an intrusion detection system using genetic algorithms. In Ossining High School Honors Thesis, 2001.

[25] Lam, Bacchus (1994). Learning bayesian belief networks: an approach based on the MDL principle. Computational Intelligence, Vol. 10, Issue No. 3, pp.269–293.

[26] E. Aleskerov, B. Freisleben, and B. Rao. Cardwatch: a neural network-based database mining system for credit card fraud detection. In Proceedings of Computational Intelligence for Financial Engineering, pages 173-200, 1997.

[27] M. Syeda, Y.-Q. Zhang, and Y. Pan. Parallel granular neural networks for fast credit card fraud detection. In Proceedings of the 2002 IEEE International Conference, volume 1, pages 572-577, 2002.

[28] R. J. Bolton and D. J. Hand. Unsupervised profiling methods for fraud detection. In conference of Credit Scoring and Credit Control VII, Edinburgh, UK, Sept 5-7, 2001.

[29] R. Brause, T. Langsdorf, and M. Hepp. Credit card fraud detection by adaptive neural data mining. In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, pages 103-106, 1999.

[30] K. C. Cox, S. G. Eick, G. J. Wills, and R. J. Brachman. Visual data mining: Recognizing telephone calling fraud. J. Data Mining and Knowledge Discovery, 1(2):225-231, 1997.

[31] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P.-N. Tan. Data mining for network intrusion detection. In Proc. NSF Workshop on Next Generation Data Mining, Baltimore, MD, November, 2002.

[32] J. R. Dorronsoro, F. Ginel, C. Sanchez, and C. S. Cruz. Neural fraud detection in credit card operationK. J. Ezawa and S. W. Norton. Constructing bayesian networks to predict uncollectible telecommunications accounts. Ieee Expert-Intelligent Systems & Their Applications, 11:45-51, 1996.s, volume 8, pages 827-834, 1997.

[33] T. F. Lunt, A. Tamaru, F. Gilham, R. Jagannathan, P. G. Neumann, H. S. Javitz, A. Valdes, and T. D. Garvey. A Real-Time Intrusion Detection Expert System (IDES) - Final Technical Report. Technical report, SRI Computer Science Laboratory, SRI International, Menlo Park, CA, Feb. 1992.

[34] D. Anderson, T. Frivold, A. Tamaru, and A. Valdes. Next-generation intrusion detection expert system (nides), software user’s manual, beta-update release. Technical Report SRI-CSL-95-07, Computer Science Laboratory, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493, May 1994.

[35] A. K. Ghosh and A. Schwartzbard. A study in using neural networks for anomaly and misuse detection. In Proceedings of the 8th USENIX Security Symposium, D.C., 1999.

[36] J. Hollmn and V. Tresp. Call-based fraud detection in mobile communication networks using a hierarchical regime-switching model. In Proceedings of the 1998 conference on Advances in neural information processing systems II, pages 889-895. MIT Press, 1999.

[37] S. Rosset, U. Murad, E. Neumann, Y. Idan, and G. Pinkas. Discovery of fraud rules for telecommunications challenges and solutions. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 409-413. ACM Press, 1999.

[38] K. Ilgun. USTAT: A real-time intrusion detection system for UNIX. In Proceedings of the 1993 IEEE Symposium on Research in Security and Privacy, pages 16-28, Oakland, CA, 1993.

[39] G. Zhao, C. Zhang and L. Zheng, Intrusion detection using deep belief network and probabilistic neural network IEEE International Conference on CSE and EUC, pp. 639-642, July 2017.

[40] N. T. Van, T. N. Thinh and L. T. Sach, an anomaly–based network intrusion detection system using deep learning IEEE International Conference on ICSSE, pp. 210-214, July 2017.

[41] T. Ishitaki, R. Obukata, T. Oda and L. Barolli, Application of deep recurrent neural networks for prediction of user behavior in tor networks IEEE 31st International Conference on WAINA, pp. 238-243, March 2017.

[42] N. Shone, T. N. Ngoc, V. D. Phai and Q. Shi, A deep learning approach to network intrusion detection IEEE International Conference on emerging topics in computational intelligence, vol. 2, pp. 41-50, January 2018.

[43] P. Vincent, H. Larochelle and I. Lajoie, stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion Journal of Machine Learning Research, Canada, pp. 3371-3408, 2010.

[44] M. Z. Alom, V. Bontupalli and T. M. Taha, Intrusion detection using deep belief networks IEEE International Conference on NAECON, pp. 339-344, June 2015.