# Intelligent and Real-Time Smart Card Fraud Detection for Optimized Industrial Decision Process

Simeon Okechukwu Ajakwe, Cosmas Ifeanyi Nwakanma, Dong-Seong Kim, Jae Min Lee
*Department of IT Convergence Engineering*
*Kumoh National Institute of Technology* Gumi, South Korea
simeonajlove@gmail.com,(cosmas.ifeanyi,dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—Smart Card fraud cases are on the rise due to increased online and real-time untact transactions triggered by several cashless policies as well as the global COVID-19 pandemic. This paper evaluates various machine learning algorithms for timely and intelligent detection and predictions of smart card frauds. The test the accuracy of the model, dataset of various frauds on smart card was used. Timely, accurate, and intelligent interception of dynamic fraud patterns is crucial to curtail loss rising from security breach. Evaluation results shows that Deep Neural Network (99.91%), Convolutional Neural Network (99.92%), Artificial Neural Network (99.93%) performed better than XGBoost (97.20%), Random Forest (96.12%), Support Vector machine (96.36%), Logistics Regression (96.34%), K-Nearest Neighbour (95.07%), and Naive Bayes (94.87%) both in accuracy and precision. The deployment this deep learning algorithm will guarantee faster smart card fraud detection, improved users' trust in smart card technology, help in fraud curtailment, facilitate timely strategic counter response measure, as well as trigger further research on improving smart card security, amongst others.**

***Index Terms*—Smart Card, Fraud, Real-Time, Deep Learning**

## I. Introduction

The unending COVID-19 global pandemic amongst other measures such as consistent push for cashless transactions, quest for creative and innovative ways of task performance, consistent drive for globalization, etc. spiked the proliferation of smart card usage both for industrial and business purposes. The ease of use, convenience, and comfort of smart card makes it attract quantum usage across the globe [1]. Due to this popularity in usage, Smart card theft and fraud has increased astronomically especially in recent times. Available statistics as seen in Fig. 1 shows that in the United States, smart card fraud grew by more than 35% between 2014 and 2018.

There is geometric increase in smart card fraud across various industrial sectors with heavy financial consequences on both the users and financial institutions [2]. The risks of carrying out industrial process or doing business with unauthorized or incorrectly identified individuals could result in financial loss, reputation damage, unauthorized access to private facilities, disclosure of confidential information, corruption of data or unenforceable agreements [3]. There are different methods of detecting smart card frauds ranging from traditional to automated methods with its attendant difficulties and complexities. One of such difficulties is evolving behavioural patterns of fraud perpetrators [4] [5]. Artificial Intelligence (AI) models

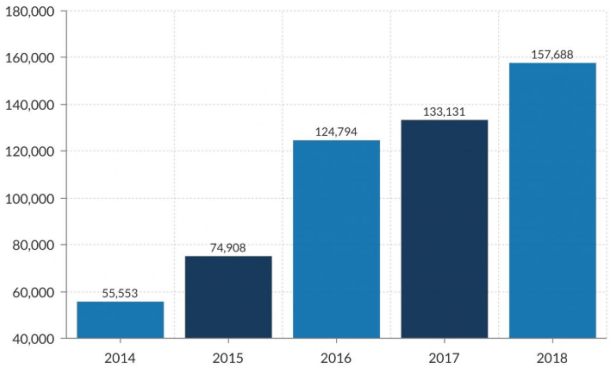


Fig. 1. Smart Card Fraud Report in US

are very useful for accurate detection and classification [6]. Several mathematical and artificial intelligent approaches have been proposed to detect smart card fraud such as data shift quantification approach, facial recognition approach, Concept-Drift Adaptation approach,HMM-based feature engineering approach, Sequence classification approach, etc [7].

However, these approaches lack real-time efficient and accurate detection based on the evolving behavioural patterns of smart card fraud perpetrators on real-time datasets from varied sources for timely counter-measure decision making which is essential in time-sensitive industrial systems as well as enterprise resource systems.

Statistical modelling of rare events using Synthetic Minority Oversampling Technique (SMOTE) is a powerful sampling method that is used to solve the imbalanced classification problem of under and over-fitting of dataset sampling. SMOTE increases the features available to each dataset class and makes the samples more general. Synthetic Minority Over-sampling Technique-Nominal Continuous (SMOTE-NE) is a modification SMOTE algorithm that handles samples with a mixture of continuous and categorical features.

In this study, SMOTE technique is applied to balance the dataset for optimized performance of selected algorithms for evaluation. The major contribution of this work is to evaluate the detection efficiency of these algorithms for timely decision making to curtail smart card fraud. Specifically, the paper examines the various smart card fraud detection detection approaches, explores their peculiarities, derived, trained and simulate datasets on machine learning algorithms to make pre-

dictions, and evaluate their performance based on fundamental metrics

The paper arrangement is as follows: Section II summarises the methodology adopted in this work. In Section III, the result and performance evaluation is presented. Paper was concluded in IV.

## II. Methodology and Data Analysis

Machine learning methodology/process is adopted in this paper which ranging from data acquisition, exploratory analysis, to eventual model evaluation and optimization. The dataset is contains 284808 samples of smart card related-frauds with 26features. Data-balancing is carried out using SMOTE technique to avoid model under/over-fitting. which mathematically is summarised as:

$$\Delta(O1,O2) = \sum_{f=0}^{F} \delta(O1_f, O2_f)^1 = \delta(A,A) + \delta(B,F) + \delta(C,C) + \delta(D,G) + \delta(E,N)$$

$$= \sum_{c=0}^{1} |p(c|A_0) - p(c|A_0)|^1 + \sum_{c=0}^{1} |p(c|B_1) - p(c|F_1)|^1 + \sum_{c=0}^{1} |p(c|C_2) - p(c|C_2)|^1$$

$$+ \sum_{c=0}^{1} |p(c|D_3) - p(c|G_3)|^1 + \sum_{c=0}^{1} |p(c|E_4) - p(c|N_4)|^1$$

where $(vf)_c$ = the number of occurrences of feature value $vf$ for class $c$, and $(vf)_D$= is the total number of occurrences of $vf$ in the dataset.

The system takes fraud pattern data as input, process it based each algorithm, and produce an output. Algorithms performance appraisal is based on accuracy and time complexity selected algorithms the assumption that algorithms with high accuracy and with least execution time are considered better than the rest for predicting smart card fraud. The accuracy is calculated by equations (1), (2) , (3), and (4) respectively.

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \quad (1)$$

$$Precision = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \quad (2)$$

$$Recall = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \quad (3)$$

$$Accuracy = \mathrm{TP} + \mathrm{TN}/(\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}) \quad (4)$$

where TP, FP and FN represent true positive, false positive and false negative, respectively.

## III. Evaluation of Results

The summary of the smart card fraud detection across the selected models are as shown in Table I, indicating the accuracy, precision, recall, f1-score, and execution time of each model. Table I shows that CNN has the highest accuracy of 99.94% with least execution time of 8s showing superior performance than DNN and ANN with 99.93% and execution time of 10s. Also, the result indicates the deep learning models outperformed the machine learning models in terms of accuracy but takes relatively more time to execute.However, the trade-off between accuracy and timeliness is subject to the decision-makers and sensitivity of the scenario at hand.

TABLE I
Results of various algorithms

| Models | F1 Score | Accuracy | Recall | Precision | Time |
|---|---|---|---|---|---|
| LR | 0.8965 | 0.9634 | 0.8666 | 0.9285 | 8s |
| KNN | 0.9176 | 0.9507 | 0.8666 | 0.9750 | 8.21s |
| SVM | 0.8863 | 0.9636 | 0.8666 | 0.9069 | 8.42s |
| NB | 0.8809 | 0.9487 | 0.8222 | 0.9487 | 8.50s |
| XGB | 0.8837 | 0.9720 | 0.8444 | 0.9268 | 8.56s |
| RF | 0.8863 | 0.9612 | 0.8666 | 0.9069 | 9.2s |
| ANN | 0.7874 | 0.9993 | 0.6802 | 0.9345 | 10s |
| ANN + SMOTE | 0.9962 | 0.9962 | 0.9995 | 0.9929 | 19s |
| CNN | 0.8181 | 0.9994 | 0.7346 | 0.9230 | 8s |
| 8CNN + SMOTE | 0.9991 | 0.9991 | 1.0000 | 0.9983 | 16s |
| DNN | 0.8040 | 0.9993 | 0.8095 | 0.7986 | 10s |
| DNN+SMOTE | 0.9991 | 0.9991 | 0.9999 | 0.9983 | 20s |

The graph in Fig. 2 shows the prediction accuracy of the compared models with CNN having the highest fraud detection accuracy and NB with the least fraud detection accuracy.

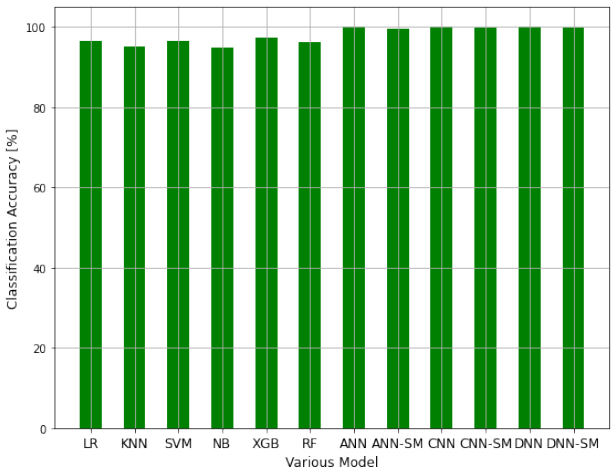


Fig. 2. Smart Card Fraud Detection Accuracy across Models

Fig. 3 and Fig. 4 shows the accuracy graph and confusion matrix of the highest performing smart card detection model,

the CNN. Note that CNN had an accuracy of 99.95% and in the confusion matrix,it is evident that it has the least misclassification of the fraud and no fraud cases in the dataset. This suggests a reasonable reduction in false alarm rate.

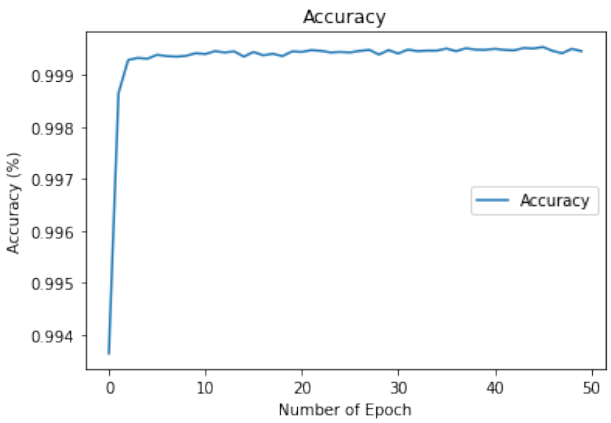


Fig. 3. Time Complexity of CNN (least time) Model

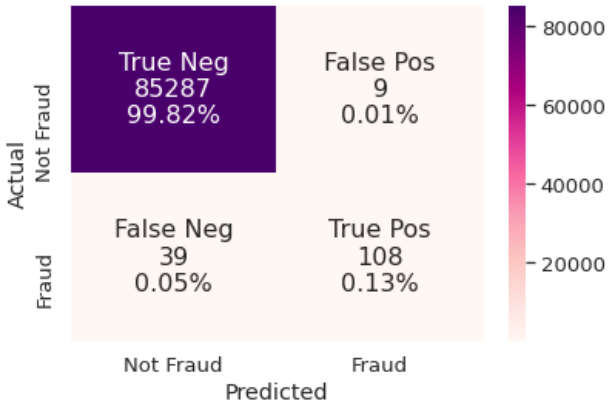


Fig. 4. Confusion Matrix of CNN (highest performing) model

## IV. Conclusion

This work evaluated the efficiency various machine learning and deep learning algorithms for detection of smart card fraud using SMOTE technique to balance dataset for optimized performance. The results reveals that DNN, CNN, and ANN performed better than RF, SVM, NB, XGB, and LR respectively with highest level of detection accuracy and precision of 99.9%, making them well suited for time-sensitive industrial process and financial decision-making for fraud curtailment amongst other applications and use-cases. Deep learning algorithms can detect and predict events in time-sensitive scenarios with optimum accuracy which is very crucial real-time systems.

## Acknowledgment

This research work was supported by Priority Research Centers Program through NRF funded by MEST(2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

## References

[1] 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.

[2] J. Jurjen and R. Leukfeldt, "Phishing and Malware Attacks on Online Banking Customers in the Netherlands: A Qualitative Analysis of Factors Leading to Victimization," *International Journal of Cyber Criminology*, vol. 10, no. 1, pp. 79–91, 2016.

[3] B. U. Islam Khan, R. F. Olanrewaju, F. Anwar, and M. Yaacob, "Offline OTP Based Solution for Secure Internet Banking Access," in *2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, 2018, pp. 167–172.

[4] Y. Lucas, P.-E. Portier, L. Laporte, S. Calabretto, L. He-Guelton, F. Oblé, and M. Granitzer, "Dataset Shift Quantification for Credit Card Fraud Detection," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2019, pp. 97–100.

[5] C. I. Nwakanma, M. S. Hossain, J.-M. Lee, and D. Kim, "Towards Machine Learning Based Analysis of Quality of User Experience (QoUE)," *International Journal of Machine Learning and Computing*, vol. 10, pp. 752–758, 2020.

[6] D. H. Kim, Y. J. Kim, and D. S. Hur, "Artificial Neural Network Based Breakwater Damage Estimation Considering Tidal Level Variation," *Ocean Engineering*, vol. 87, pp. 185–190, 2014.

[7] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection and Concept-Drift Adaptation with Delayed Supervised Information," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–8.