

Credit card fraud detection using CNN and LSTM

Nishant Upadhyay¹, Nidhi Bansal², Divya Rastogi³, Rekha Chaturvedi⁴, Mohammad Asim¹,
Suraj Malik⁵, Khel Prakash Jayant⁶, Abhay Kumar Vajpayee⁷

¹Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India

²Department of Computer Science and Engineering, School of Engineering and Technology, Manav Rachna International Institute of Research and Studies (Deemed to be University), Faridabad, Haryana, India

³Department of Computer Science and Application, School of Engineering and Technology, Sharda University, Greater Noida, India

⁴Department of Data Science and Engineering, School of Information Security and Data Science, Manipal University Jaipur, Jaipur Rajasthan, India

⁵Department of Computer Science and Engineering, IIMT University, Meerut, India

⁶Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India

⁷Department of Computer Science and Engineering, Institute of Engineering and Technology Sitapur, Lucknow, India

Article Info

Article history:

Received May 13, 2024

Revised Oct 10, 2024

Accepted Oct 30, 2024

Keywords:

CNN

Credit card

Fraud detection

LSTM

Online transaction

ABSTRACT

Credit card fraud is an evolving problem with the fraudsters developing new technologies to perform fraud. Fraudsters have found diverse ways to make a fraud transaction to the card holder. Thus, detecting suspicious behavior of a card is critical for preventing fraudulent transactions to happen. Artificial intelligence techniques, in particular deep learning algorithms can tackle these credit card fraud attacks by identifying patterns that predict transactions as fraud or legitimate. One-dimensional convolutional neural network (1D CNN) and long short-term memory (LSTM) both performs well on the sequential data especially on transactions data, yet there are not many studies done on combining these two algorithms to make an effective fraud detection approach. However, the dataset is highly imbalanced containing only 492 fraud transaction out of two lacs transactions. In this experimental study, firstly datasets will get prepared by using different sampling techniques along with their hybrid techniques secondly, observing the performance of individual CNN and LSTM on the datasets, finally on those datasets in which CNN and LSTM are performing well, by implementing ensemble on those data. The performance of the ensembles is observed using the performance metrics namely accuracy, F1-score, precision and recall. In the proposed experimental study, getting the F1-score of 99.96% and 99.89% in ensemble: early fusion and ensemble: late fusion respectively.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rekha Chaturvedi

Department of Data Science and Engineering, School of Information Security and Data Science

Manipal University Jaipur

Jaipur Rajasthan, India

Email: rekha.chaturvedi@jaipur.manipal.edu

1. INTRODUCTION

With the digital transactions, modern banking is getting smart transactions over the internet. However, this ease of use has attracted malicious actors and increased concerns about credit card fraud. The potential financial losses are significant, with reports indicating billions of euros lost annually in Europe alone [1]. Fraudsters exploit various tactics, from compromising data on public Wi-Fi to utilizing underground marketplaces [2]. While researchers have developed numerous fraud detection methodologies,

they often struggle with high false positive rates and difficulty adapting to evolving fraud patterns [3], [4]. Additionally, traditional machine learning approaches raise concerns about data privacy. Deep learning algorithms such as convolutional neural networks (CNN) and long short-term memory (LSTM), have emerged as effective tools for detecting fraud by identifying intricate patterns and behaviors in transaction data. However, due to highly imbalanced nature of the credit card data, implementing these algorithms poses a significant challenge.

While, individual studies have shown promising results in fraud detection using CNNs and LSTMs, there is limited research on combining these two algorithms to build a powerful ensemble for fraud detection [5]. This experimental study aims to address this gap by building an efficient ensemble through ensemble through early and late fusions of CNNs and LSTMs. To account for the imbalanced dataset, various sampling techniques are incorporated, including hybrid sampling methods, to evaluate model performance under different circumstances.

2. METHOD

The proposed method and implementation include various key points which are listed here capable of making the system efficient for useful transactions. Executive summary: credit card fraud is an evolving problem which can cost businesses and people money. This study investigates the viability of detecting credit card fraud using an ensemble model that combines LSTM with CNNs [6], [7]. This strategy may increase the accuracy of fraud detection by utilizing the advantages of both CNN and LSTM in processing sequential data and collecting spatial information. Project description: the projects objective is to use a CNN-LSTM ensemble model to design and assess a credit card fraud detection system. Credit card transaction data, comprising sequential (such as a transaction history) and static (such as cardholders' details and location) information, will be processed by the system. Market analysis: financial institutions such as banks, credit card companies are looking to enhance their fraud detection skills are part of the target market. The global fraud losses are expected to reach \$206 billion by 2025, indicating the scale of this market [8], [9]. Current fraud detection programs provided by security firms and financial institutions themselves are competitors. By combining the benefits of both CNNs and LSTMs, our suggested ensemble model may be advantageous in terms of increased detection rate. Technical feasibility includes:

- Strengths: CNNs are particularly good at removing geographical characteristics, such as location, cardholder details from data. Transaction history is one type of sequential data that LSTMs are good at capturing temporal trends in. Fusing both models through ensemble learning may result in improved performance.
- Challenges: due to the model's intricacy, training will take a large amount of processing power. It can take a while to fine-tune the hyper-parameters for the CNN and LSTM components. For training to be effective, a sizeable, labeled credit card transaction dataset must be available.
- Technical assets: a number of open-source frameworks, such as PyTorch and TensorFlow, can make a model development easier. Platform for cloud computing provide scalable resources for sophisticated model training.

Figure 1 is able to show the flow of work or a plan for executing a task. Where the data set needs to be fetched first in order to get it into preprocessing [10], [11]. This will further move towards necessary implementation. And finally, the system will be able to show the results. CNN architectures that were created to handle the qualities of the dataset were used in the methodology used in this work. This approach established a foundational benchmark for the evaluation of more complex architectures. Deep CNNs: commonly referred to as CNNs, ConvNets, or DCNNs, are in the fields of computer vision and image processing because of their ability to interpret data in the form of many arrays.

As seen in Figure 2, the first layer, which is often a convolutional layer, uses a set of mathematical operations to identify features including edges, textures, and shapes. Subsequently, the pooling layers reduce the spatial dimensions of the representation, thereby reducing the number of parameters and calculations within the network. The network usually consists of fully connected layers after several convolutional and pooling layers. These layers are typical neural network layers in which a learned weight connects each input to each output. To categorize or predict the output at this point, the network integrates all of the features that it has learned from the earlier layers.

2.1. Dataset

In September 2013, European cardholders conducted credit card transactions that are included in the databases. There are 492 frauds out of 284,807 transactions in this dataset shown by Figures 3 and 4. The graph above demonstrates that the two most popular transaction methods are TRANSFER and CASH_OUT. It also demonstrates that fraud can only occur through these two methods. The model has identified false

positives but never let even a single false negative which is more important than FP. Since we can't miss out a fraud transaction, but we can manage false positive results by investigating them.

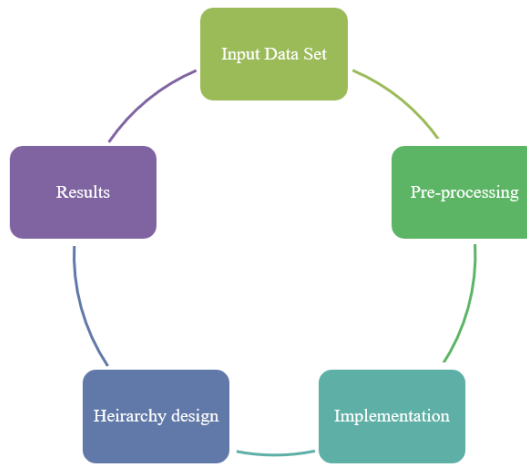


Figure 1. Plan for execution

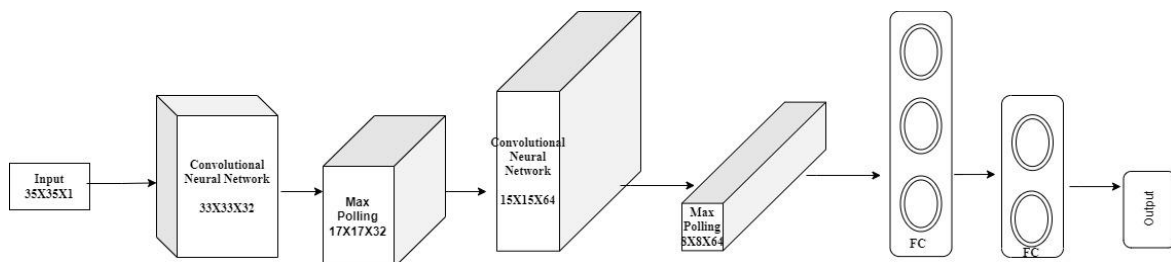


Figure 2. Deep convolutional neural networks

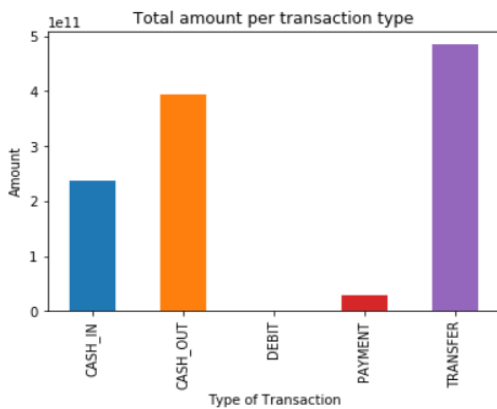


Figure 3. Type of transaction

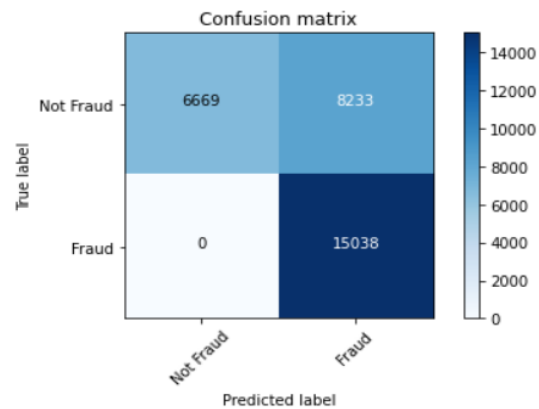


Figure 4. Confusion matrix

Experimental study adopted a deep learning approach to tackle credit card fraud detection. Here focused for two prominent models - CNNs and LSTMs. Then we'll be building an ensemble of these models, namely ensemble early fusion: CNN-LSTM and ensemble late fusion: CNN-LSTM [12]. These models were chosen for their ability to learn complex patterns within credit card transactions dataset. To prepare the data for analysis from the highly imbalanced credit card fraud dataset implemented the series of preprocessing steps [13]. This includes main tasks like standardization, reshaping and then resampling. Following the data

preprocessing stage, later designed separate architectures for both CNN and LSTM. These architectures define the structure of the models, including the types of layers used, their activation functions [14]. Assess the performance of these models, utilized various evaluation metrics namely accuracy, recall, precision and F1-score. Finally, building ensembles namely ensemble early fusion: CNN-LSTM and ensemble late fusion: CNN-LSTM and then observing their performance using the performance metrics.

2.2. Test cases

Here how the data was split into training, validation and testing sets, will get exploration. The training set is used to train the models, the validation set is used for hyper parameter tuning, and the testing set provides an independent measure of the model performance in the unseen data. Employed an 80-20 train-test split strategy to divide our credit card transaction data. Here, 80% of the data was allocated for training the models, allowing them to learn the patterns within legitimate and fraudulent transactions. The remaining 20% of the data was designated as the test set [15]. It is important to note that this 20% test set was further divided into a validation and a final testing set. A small portion of the initial 20% test data was used as the validation test [16]. This validation set played a crucial role in hyperparameter tuning. By evaluating models' performance on the validation set during training, we could adjust hyper parameters like number of epochs or learning rate, to optimize the model's performance without overfitting on the final testing set.

The remaining portion of the initial 20% test data served as the final testing set, also shown by Figures 5 and 6. This unseen data provided a more objective evaluation of the model's ability to generalize to real world scenarios. All models individual CNN, LSTM and their ensembles namely ensemble early fusion and ensemble late fusion were evaluated on the final testing set using the various performance metrics like accuracy, precision, recall, and F1-score.

```
print("----- Splitting Datasets -----")
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.2)
print("----- Splitted Successfully -----")
```

Figure 5. Train-test sets

```
print("----- Model Fitting -----")
history = model_cnn.fit(X_train, y_train, epochs=num_epochs, validation_data=(X_test, y_test), verbose=0, callbacks=[es, mc])
print("----- Fitted Successfully -----")
```

Figure 6. Validation sets

3. RESULTS AND DISCUSSION

By proposing and designing the model as per the theme of the proposed working system the efficiency is being increased and also showing better performance when comparing with various other models. Moreover, test cases are also listed in the paper to secure the code within the execution phase, and finally reach the optimal solution in terms of improved accuracy and efficiency of the system. With deep learning approach: due to the continuous nature outputs of the predict function, converted it into binary i.e 0 or 1. By performing this conversion at three different thresholds that is at 0.5, 0.6 and 0.7. Out of which results in better performance, has been taken into account [17], [18]. According to numerical valued results, that must be only on synthetic minority over-sampling technique (SMOTE) data, Near Miss Under-sampling (NMUS) data, over sampled (OS) data, hybrid: OS-NMUS data and hybrid: SMOTE-NMUS data both models are showing exceptional results. Worst performance of CNN and LSTM has been observed in original and scaled data. OS data shows better performance of CNN and LSTM model are noted at thresholds 0.7 and 0.5 respectively. In this case CNN outperforms LSTM with an accuracy and F1-score of 99.89% and 99.89% respectively, with 100% recall [19].

And also, can say that that CNN is properly classifying 56,916 transactions as fraud, 56,687 transactions as legit, 123 legit transactions as fraud and 0 fraud transactions as legit. However, LSTM is properly classifying 55,224 transactions as fraud, 55,698 transactions as legit, 1,240 legit transactions as fraud and 1,564 fraud transactions as legit. Figures 7 to 16 are depicting the actual results along with the assumed test cases. SMOTE shows better performance of CNN and LSTM model are noted at thresholds 0.7 and 0.5 respectively, where CNN again outperform LSTM with an accuracy and F1-score of 99.92% and 99.92% respectively, with 100% recall. And also, can say that CNN is properly classifying 56,757 transactions as fraud, 58,881 transactions as legit, 88 legit transactions as fraud and 0 fraud transactions as legit. However, LSTM is properly classifying 53,471 transactions as fraud, 55,962 transactions as legit, 1,006

legit transactions as fraud and 3,282 fraud transactions as legit. NMUS shows CNN outperforms LSTM as in case of CNN, getting all accuracy, F1-score, recall, and precision of 100%. And also, can say that CNN is properly classifying 92 transactions as fraud, 105 transactions as legit, 0 legit transactions as fraud and 0 fraud transactions as legit [20]. However, LSTM is properly classifying 94 transactions as fraud, 116 transactions as legit, 0 legit transactions as fraud and 2 fraud transactions as legit. On hybrid: OS-NMUS demonstrates the performance of the models on hybrid sampling dataset of OS-NMUS. Better performance of CNN and LSTM is noted at thresholds 0.7 and 0.5 respectively. Clearly CNN outperforms LSTM with an accuracy and F1-score of 99.89% and 99.84% respectively with 100% recall [21], [22]. And also, can say that CNN is properly classifying 5,734 transactions as fraud, 11,307 transactions as legit, 18 legit transactions as fraud and 0 fraud transactions as legit. However, LSTM is properly classifying 5,468 transactions as fraud, 11,359 transactions as legit, 87 legit transactions as fraud and 145 fraud transactions as legit. SMOTE-NMUS data demonstrates the performance of the models on hybrid sampling dataset of OS-NMUS. Better performance of CNN and LSTM is noted at thresholds 0.7 and 0.5 respectively, where CNN outperforms LSTM with an accuracy and F1-score of 99.86% and 99.79% respectively. And also, can say that CNN is properly classifying 5,736 transactions as fraud, 11,300 transactions as legit, 22 legit transactions as fraud and 1 fraud transactions as legit. However, LSTM is properly classifying 5,287 transactions as fraud, 11,377 transactions as legit, 70 legit transactions as fraud and 325 fraud transactions as legit.

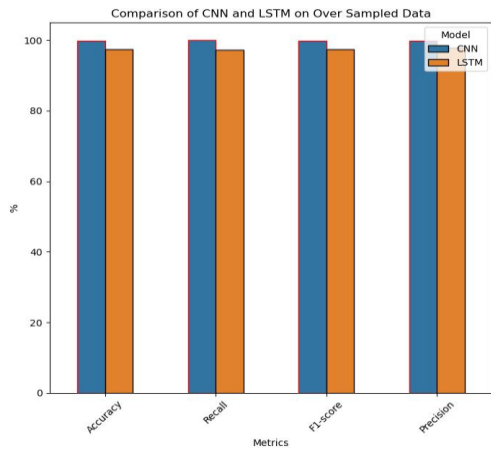


Figure 7. CNN/LSTM results (sample data)

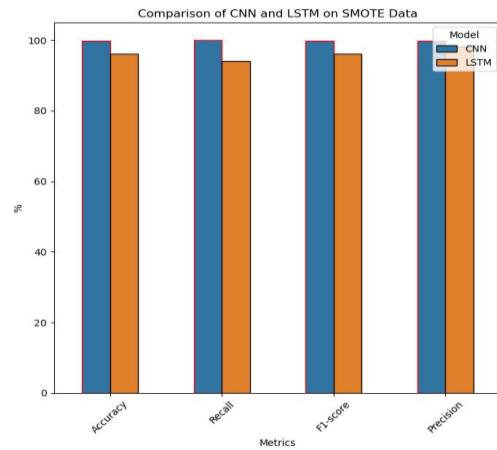


Figure 8. CNN/LSTM results (SMOTE data)

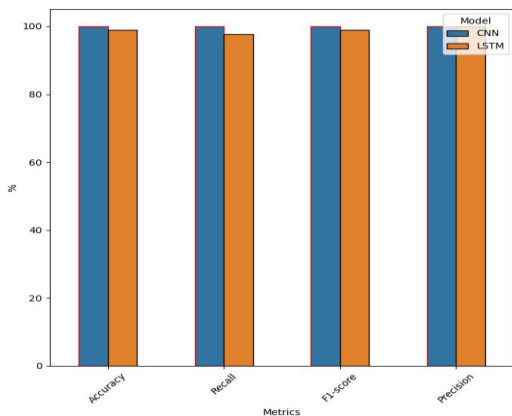


Figure 9. CNN/LSTM results (NMUS)

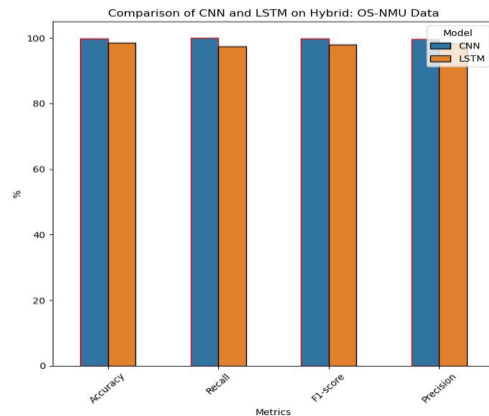


Figure 10. CNN/LSTM results (hybrid)

With ensemble approaches: the performance of two ensemble learning models namely early fusion: CNN-LSTM and late fusion. It shows that the early fusion is resulting better on the datasets compared to late fusion. Below will be doing through analysis of the impact of these datasets on the performance of ensembles. NMUS data shows the performance of the ensembles, on the NMUS dataset.

Threshold=0.5, 0.6, 0.7 giving the same result in case of early fusion and threshold=0.5 in late fusion. Both models have high accuracy, the late fusion outperforms the early fusion, as it outputs high accuracy and F1-score of 100% and 100% respectively [23]. And also, can say that early fusion is properly classifying 96 transactions as fraud, 99 transactions as legit, 0 legit transactions as fraud and 2 fraud transactions as legit. However, late fusion is properly classifying 98 transactions as fraud, 99 transactions as legit, 0 legit transactions as fraud and 0 fraud transactions as legit. On SMOTE demonstrates the performance of the ensembles, on the SMOTE dataset. Threshold=0.7 giving the best result in case of early fusion and threshold=0.5 in late fusion. Early fusion outperforms the late fusion, as it outputs high accuracy and F1-score of 99.96% and 99.96% respectively. And also, can say that early fusion is properly classifying 56,976 transactions as fraud, 56,710 transactions as legit, 40 legit transactions as fraud and 0 fraud transactions as legit [24]. However, late fusion is properly classifying 56,838 transactions as fraud, 56,641 transactions as legit, 109 legit transactions as fraud and 138 fraud transactions as legit. Over OS demonstrates the performance of the ensembles, on the OS dataset. Threshold=0.7 giving the best result in case of early fusion and late fusion. Late fusion outperforms the early fusion, as it outputs high accuracy and F1-score of 99.89% and 99.89% respectively.

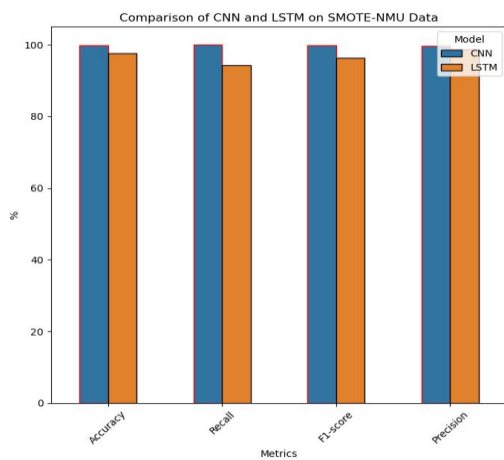


Figure 11. CNN/LSTM results (SMOTE-NMU)

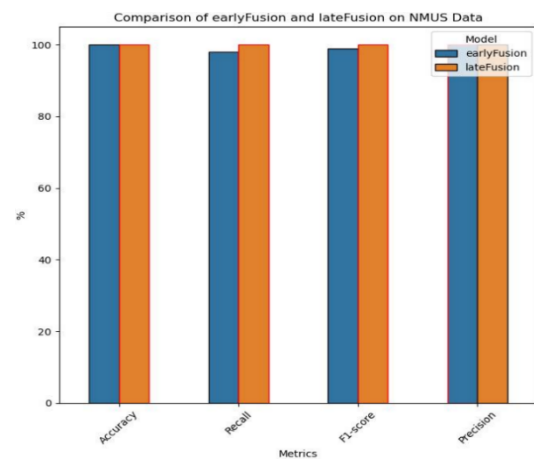


Figure 12. CNN/LSTM results (NMUS-fusion)

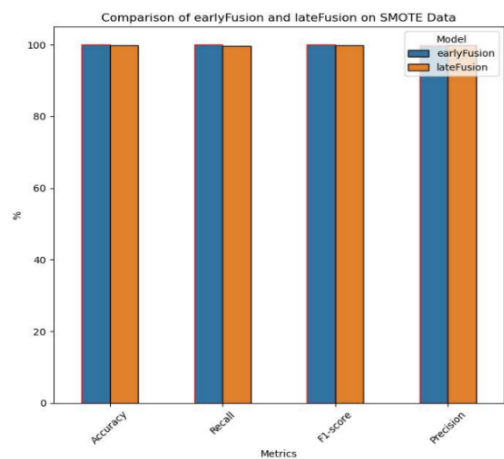


Figure 13. CNN/LSTM results (SMOTE-fusion)

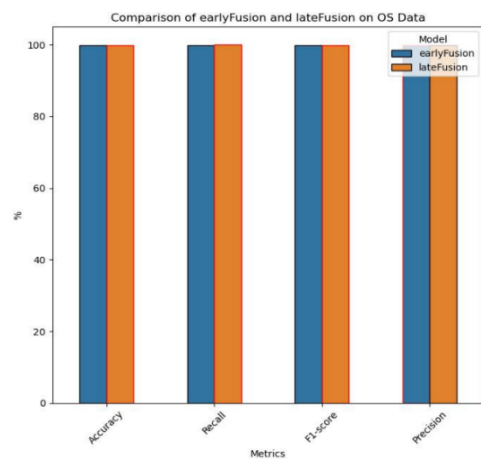


Figure 14. CNN/LSTM results (OS data)

And also, can say that early fusion is properly classifying 56,637 transactions as fraud, 56,923 transactions as legit, 57 legit transactions as fraud and 109 fraud transactions as legit. However, late fusion is properly classifying 56,746 transactions as fraud, 56,862 transactions as legit, 118 legit transactions as fraud and 0 fraud transactions as legit. On hybrid: (OS-NMUS) data demonstrates the performance of the ensembles, on the OS-NMUS dataset. Threshold=0.7 giving the best result in case of early fusion and threshold=0.5 in late fusion. Early fusion outperforms the late fusion, as it outputs high accuracy and F1-score of 99.91% and 99.86% respectively [25]. And, can say that early fusion is properly classifying 5,731

transactions as fraud, 11,313 transactions as legit, 8 legit transactions as fraud and 7 fraud transactions as legit. However, late fusion is properly classifying 56,660 transactions as fraud, 56,561 transactions as legit, 189 legit transactions as fraud and 316 fraud transactions as legit. SMOTE-NMUS data demonstrates the performance of the ensembles, on the SMOTE-NMUS dataset. Threshold=0.5 giving the best result in case of early fusion and in late fusion. Early fusion outperforms the late fusion, as it outputs high accuracy and F1-score of 99.94% and 99.92% respectively. And, can say that early fusion is properly classifying 5,734 transactions as fraud, 11,316 transactions as legit, 5 legit transactions as fraud and 4 fraud transactions as legit. However, late fusion is properly classifying 5,664 transactions as fraud, 11,311 transactions as legit, 10 legit transactions as fraud and 74 fraud transactions as legit.

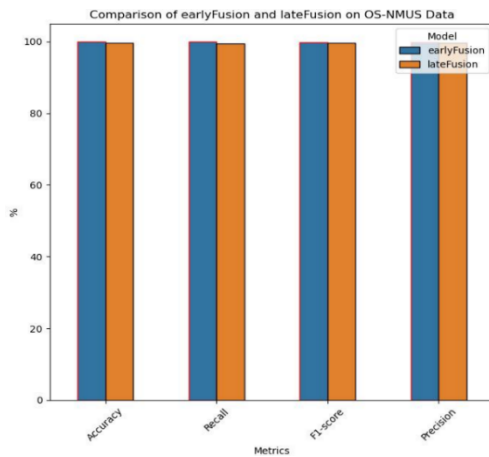


Figure 15. CNN/LSTM results (OS-NMUS)

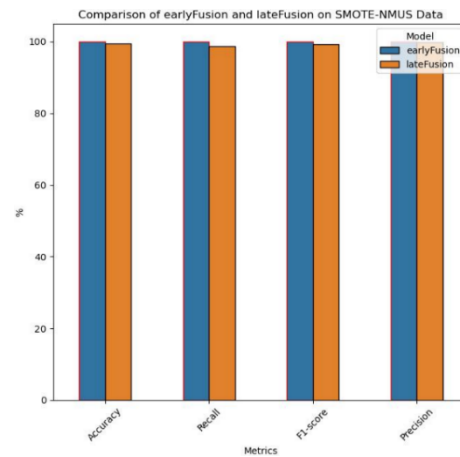


Figure 16. CNN/LSTM results (SMOTE-NMUS)

4. CONCLUSION

From the analysis, across various datasets, both models CNN and LSTM shown impressive results, generally achieving high accuracy and F1-scores. However, some key differences exist. Dataset generated using NMUS, an under-sampling technique and SMOTE, an oversampling technique show great accuracy and F1-scores, suggesting models are doing well on these datasets. Original and scaled datasets, on the other hand, exhibit lower performance, particularly for LSTMs, hinting at imbalanced data or inherent challenges. Encouragingly ensemble models often surpass individual models, demonstrating the benefits of combining diverse learning styles. Early fusion ensembles typically edge out late fusion approach, suggesting that fusing features before individual model predictions are more effective, especially in our case. Hybrid datasets, combining SMOTE and NMUS, see the most significant gains from the ensemble models, even reaching near-perfect accuracy in some instances.





REFERENCES

- [1] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020, doi: 10.1016/j.ins.2019.11.004.
- [2] D. Dhiman, A. Bisht, A. Kumari, D. H. Anandaram, S. Saxena, and K. Joshi, "Online fraud detection using machine learning," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Jan. 2023, pp. 161–164, doi: 10.1109/AISC56616.2023.10085493.
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [4] A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, pp. 18–25, 2018, doi: 10.14569/IJACSA.2018.090103.
- [5] F. M. Moreno, J. A. N. Andrade, H. T. Sinohara, and P. H. H. N. de Araujo, "Transfer learning," *Journal of critical reviews*, vol. 7, no. 14, 2020, doi: 10.31838/jcr.07.14.174.
- [6] S. Maes, K. Tuyls, and B. Vanschoenwinkel, "Credit card fraud detection using Bayesian and neural networks," *Maciunas RJ, editor. Interactive image-guided neurosurgery. American Association Neurological Surgeons*, pp. 261–270, 1993.
- [7] Z. Chen, S. Wang, D. Yan, and Y. Li, "Research and implementation of bank credit card fraud detection system based on reinforcement learning and LSTM," in *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Dec. 2023, pp. 1–8, doi: 10.1109/ICMNWC60182.2023.10435890.
- [8] M. L. Gambo, A. Zainal, and M. N. Kassim, "A convolutional neural network model for credit card fraud detection," in *2022 International Conference on Data Science and Its Applications (ICoDSA)*, Jul. 2022, pp. 198–202, doi: 10.1109/ICoDSA55874.2022.9862930.





- [9] Y. A. Mohmad, "Credit card fraud detection using LSTM algorithm," *Wasit Journal of Computer and Mathematics Science*, vol. 1, no. 3, pp. 26–35, 2022, doi: 10.31185/wjcm.60.
- [10] B. P. Verma, V. Verma, and A. Badholia, "Hyper-tuned ensemble machine learning model for credit card fraud detection," in *2022 International Conference on Inventive Computation Technologies (ICICT)*, Jul. 2022, pp. 320–327, doi: 10.1109/ICICT54344.2022.9850940.
- [11] D. Kaur, A. Saini, and D. Gupta, "Credit card fraud detection using machine learning, deep learning, and ensemble of the both," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Nov. 2022, pp. 484–489, doi: 10.1109/PDGC56933.2022.10053175.
- [12] J. Karthika and A. Senthilselvi, "Credit card fraud detection based on ensemble machine learning classifiers," in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Aug. 2022, pp. 1604–1610, doi: 10.1109/ICESC54411.2022.9885649.
- [13] P. Tomar, S. Shrivastava, and U. Thakar, "Ensemble learning based credit card fraud detection system," in *2021 5th Conference on Information and Communication Technology (CICT)*, Dec. 2021, pp. 1–5, doi: 10.1109/CICT53865.2020.9672426.
- [14] G. Gursoy and A. Varol, "Risks of digital transformation: review of machine learning algorithms in credit card fraud detection," in *2021 2nd International Informatics and Software Engineering Conference (IISEC)*, Dec. 2021, pp. 1–6, doi: 10.1109/IISEC54230.2021.9672354.
- [15] G. K. Arun and K. Venkatachalapathy, "Convolutional long short term memory model for credit card detection," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Nov. 2020, pp. 1168–1172, doi: 10.1109/ICECA49313.2020.9297606.
- [16] 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, May 2021, doi: 10.1016/j.ins.2019.05.042.
- [17] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93010–93022, 2019, doi: 10.1109/ACCESS.2019.2927266.
- [18] E. Btoush, X. Zhou, R. Gururajan, K. Chan, and X. Tao, "A survey on credit card fraud detection techniques in banking industry for cyber security," in *2021 8th International Conference on Behavioral and Social Computing (BESC)*, Oct. 2021, pp. 1–7, doi: 10.1109/BESC53957.2021.9635559.
- [19] J. H. Victoria and A. Jim, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85–126, 2004.
- [20] T. Ma *et al.*, "An unsupervised incremental virtual learning method for financial fraud detection," in *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, Nov. 2019, vol. 2019-Novem, pp. 1–6, doi: 10.1109/AICCSA47632.2019.9035259.
- [21] M. Puh and L. Brkic, "Detecting credit card fraud using selected machine learning algorithms," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2019, pp. 1250–1255, doi: 10.23919/MIPRO.2019.8757212.
- [22] N. F. Ryman-Tubb, P. Krause, and W. Gam, "How artificial intelligence and machine learning research impacts payment card fraud detection: a survey and industry benchmark," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 130–157, 2018, doi: 10.1016/j.engappai.2018.07.008.
- [23] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, Mar. 2018, pp. 1–6, doi: 10.1109/ICNSC.2018.8361343.
- [24] D. Huang, D. Mu, L. Yang, and X. Cai, "CoDetect: financial fraud detection with anomaly feature detection," *IEEE Access*, vol. 6, pp. 19161–19174, 2018, doi: 10.1109/ACCESS.2018.2816564.
- [25] T. Amarasinghe, A. Aponso, and N. Krishnarajah, "Critical analysis of machine learning based approaches for fraud detection in financial transactions," in *ACM International Conference Proceeding Series*, 2018, pp. 12–17, doi: 10.1145/3231884.3231894.

BIOGRAPHIES OF AUTHORS






Nishant Upadhyay     is an assistant professor with over five years of experience in higher education, specializing in operating systems, data structures, and machine learning. He is currently pursuing a Ph.D. and holds an M.Tech. in Data Science from Jawaharlal Nehru University (JNU). He has published research in areas such as artificial intelligence, network security, and healthcare, and holds patents for several innovations, including a smartwatch application for real-time electrical signal detection. He has taught at institutions such as GNIoT, G.B. Degree College, and is currently working at Sharda University, Greater Noida. He has also worked on research projects related to COVID-19 and clinical trials. His technical expertise includes Python, AI/ML, data analysis, and neural networks. He can be contacted at email: nishant.upadhyay23@gmail.com.






Dr. Nidhi Bansal     is an associate professor at MRIIRS Faridabad Haryana, India. She received a B.Tech. from GBTU-UPTU Lucknow India in 2010, M.E. from NITTTR Panjab University Chandigarh India in 2014, and Ph.D. in Computer Science from AKTU-UPTU Lucknow India in 2023. Her research interests are in cloud computing and machine learning broadly, with applications in data science, and computer networking. She can be contacted at email: nidhi18jul@gmail.com.






Divya Rastogi    is working with Sharda University greater Noida as assistant professor. She is MCA from UPTU Lucknow and M.Tech. (CSE) from Amity University, Noida. Now, she is pursuing a Ph.D. (CSE) from Amity University, Greater Noida Campus. Her interest area includes cloud computing, IoT, nature inspired algorithms, and image processing. She can be contacted at email: divya.rastogi03@rediffmail.com.






Dr. Rekha Chaturvedi    is currently working as assistant professor at Manipal University Jaipur. She did B.E. (IT) from Rajasthan University, M. Tech (Software Engineering) from the SGVU and Ph.D. (CSE) from Amity University Rajasthan. Her research interest includes data mining, image processing, digital image watermarking, machine learning, soft computing, and nature inspired computing. She can be contacted at email: rekhachaturvedi12@gmail.com.






Mohammad Asim    is presently working as an assistant professor in the Department of Computer Science and Engineering at Sharda University. He has vast experience in teaching and research. He did his M.Tech. (CSE) and B.Tech. (CSE) from Dr. A.P.J Abdul Kalam Technical University, Lucknow (Formerly, UPTU, Lucknow) and pursuing Ph.D. in Computer Science and Engineering. He can be contacted at email: er.mohdasim@gmail.com.






Dr. Suraj Malik    is currently working in IIMT University. Networking is considered as his research area. Ph.D. done from AKTU (UPTU Lucknow) in 2021. Many reputed publications have been done related to computer networks. He can be contacted at email: surajmalik@iimtindia.net.



Dr. Khel Prakash Jayant    teaches at RKGIT, Ghaziabad, India. He is PhD in CSE from SJPJT University, MTech in CSE, MCA from MNNIT Allahabad. Also, 24 years of experience in academics. He is also working as Former associate professor and research, EIT, Eritrea, North East Africa. His research area is AI, ML, and algorithms. He can be contacted at email: kpjayant@gmail.com.



Abhay Kumar Vajpayee    is pursuing his Ph.D. in Computer Science and Engineering from Institute of Engineering and Technology, Lucknow, UP, India. He has published many papers and given talk on various technological topics. He can be contacted at email: gyanifive@gmail.com.