Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

AI-Driven Fraud Detection in Banking: Enhancing Transaction Security

R Sabareesh,

Principal Consultant, Fortuna Business Solutions, Coimbatore, Tamilnadu, India. fortunainbox@gmail.com

Dr Deependra Nath Pathak,

Associate Professor (Law), Amity Law School, Amity University, Patna, India. dnpathak@ptn.amity.edu

Dr Radha Ranjan,

Assistant Professor (Law), Amity Law School, Amity University, Patna, India. rranjan@ptn.amity.edu

Dr Ramachandran Devi Prasanna,

Assistant Professor, Department of Commerce, Nirmala College, Coimbatore, Tamilnadu, India. fortunainbox@gmail.com

Dr. P. Shalini,

Assistant Professor, Department of Master of Business Administration, Panimalar Engineering College, Chennai, Tamilnadu, India. shalini.mbapec@gmail.com

Eijaz Khan Bellary,

Research Scholar, Department of Management Studies, Pondicherry Central Government University, Puducherry India. eijazkhan42@gmail.com

Abstract –An important step forward in risk management and fraud detection has been achieved with the integration of Artificial Intelligence (AI) in the banking sector. In this paper, we take a look at how AI has revolutionized various fields, shedding light on the benefits and drawbacks of this technology. The effects of AI on risk management are complex. More complex credit risk assessment models are made possible by algorithms that can see patterns in massive datasets that people might miss. When it comes to market and liquidity issues, real-time transaction monitoring is absolutely essential for quick risk mitigation. Automating compliance with regulatory norms is another critical function of AI, which helps to decrease human mistake and assures quick adaptability to changes in regulations. The automation of mundane processes and the reinforcement of cybersecurity measures further reduce operational risks. By examining client behaviour and transaction data, enhanced algorithms may adeptly spot anomalies that could indicate fraud. Artificial intelligence's capacity to foresee future events enables it to foil possible fraud attempts. The systems are designed to respond to changing fraudster strategies with its adaptive learning feature.

***Keywords*—**Support Vector Machine (SVM), Fraud Detection in Banking, K-Nearest Neighbor (KNN).

I. INTRODUCTION

If financial fraud increases in frequency, it might have serious consequences for the banking sector, corporations, and the government. Our reliance on web technology has considerably improved banking transactions in today's climate. However, the rise of both online and offline transactions has coincided with an increase in banking sector fraud. The sudden ascent to popularity of transactions as a payment channel has led to a refocus on procedure strategies to counteract the fraud disadvantage. Companies in various industries have invested much in software systems that can identify and prevent fraud. This includes those dealing with online shopping, credit cards, insurance, and retail. Data

mining is one of the most well-known and extensively utilized methods for detecting fraud in the financial industry. The real reasons for an application or a transaction are always a mystery[1]. The best course of action is to use mathematical algorithms to scour the accessible data for signs of fraud. Banks are facing threats to their security that are prompting them to suffer provocations. Bank security isn't precisely a customer service priority, but I think we can all agree that it could need some work. In addition, all fractals employ sophisticated analytical models and techniques to reliably detect suspicious activity. Detecting fraud requires looking for obvious indicators of dishonesty when no prior suspicion or tendency toward deception exists. Some examples of fraud include accounting, insurance, and credit card fraud[2]. The worldwide financial crisis and the subsequent regulations and sanctions have had a profound impact on banking risk management in the past ten years. However, there are already significant changes happening that will lead to an even more dramatic shift in risk management in the coming decade. As new technology improvements allow for new risk-management tactics, the risk function can make better risk decisions at lower costs. Big data, machine learning, and crowdsourcing are a few examples of potential consequences. A major issue for financial institutions is gaining customer and regulatory support for models that use online behavior and social data. When someone or some group takes advantage of loopholes in the financial system to enrich themselves or their business, this is called fraud. Businesses in today's cutthroat business climate face the formidable threat of fraud. There has been a recent escalation in this issue. The annual loss to owners of banks and other financial institutions due to fraud has skyrocketed in the last several years, reaching billions of dollars[3]. Technology improvements in various sectors are leading to the production of large amounts of data. An increasing number of complicated interrelationships is positively correlated with an increasing amount of data. In this setting, data mining refers to an exploratory data analysis technique that uses tools and techniques from a variety of scientific fields to discover hidden patterns in massive datasets. It becomes more difficult and time-consuming to get concealed information due to big data. Data mining, which make use of tools like databases, AI, and machine learning, has grown in popularity as a scientific discipline for discovering patterns in this data.

II. LITERATURE SURVEY

A number of methods have been proposed by researchers to detect and prevent financial transaction fraud. [4]discovered a novel model called the AFDM, which is based on artificial intelligence. The immune system is put to use Artificial Intelligence Revealed System (AIRS) to improve the accuracy of fraud detection. [5]investigate the feasibility of using ML algorithms to identify instances of credit card fraud. First up, the Decision trees, stochastic forests, neural networks, and Bayes regression using logistic and linear models assessment of the conventional models for support vector machines owned by no one. Using an oversampling method and three separate dataset ratios, [6] tackle the problem of data imbalance. The authors utilize three ML methods: logistic via the use of regression, K-nearest neighbor, and Naive Bayes. The person who algorithms are assessed based on how well they work, area under the curve, specificity, accuracy, F1-score, and sensitivity curve. They also show that the logistic regression model does one of the most widely used algorithms for detecting fraud[7] as stated in the text. The writers offer a system that combines the promise and cost of meta-learning ensemble methods a model for sensitive learning applicable to the detection of fraud. The smart approach suggested by [8] makes it easy to detect credit card fraud. They provide a hyperparameter optimization approach based on Bayesian principles to help with LightGBM parameter tuning. The datasets used for their experiments are publicly available credit card transaction datasets. According to this research, fraudsters are losing millions of dollars because they use several illicit ways to avoid security checks. For smaller datasets, SVM, Random Forest, and KNN can improve results [9]. For larger datasets, CNN and SVM are the best options. the third Fraud detection processes are critical for maintaining client goodwill with the firm. In this work, we provide a predictive classification model that is a combination of several popular methods, including Bagging, Extreme Learning Machine, Random Forest, Multi-layer, and KNN.[10] Credit cards have become more common as a means of payment. The expansion of credit card transactions has been accompanied by an upsurge in fraudulent activities. Customer or client payment details so that the deals can be finalized. Because of this, the store owner can't is able to verify the identity of the cardholder. To make it more sensitive, the forest model is recommended precision, individuality, and accuracy in detecting fraudulent transactions [11]. An essential banking function is the use of credit cards. Banks and other financial organizations can assess risk by looking at their activities. Random forest, the proposed algorithm, is selected due to its accuracy and effectiveness. Although SVM is a versatile method, The problem with its biased data collecting calls for additional pre- processing of data[12]. Machine learning and data analytics have revolutionized the process of identifying fraudulent activities. According to [13], there have been encouraging outcomes when ML approaches such as decision

trees, random forests, neural networks, and support vector machines are used to identify fraudulent transactions. Supervised learning models are trained to use labeled datasets to differentiate between valid and fraudulent financial transactions. As an example, logistic regression models have the ability to utilize historical data in order to predict the probability of a fraudulent transaction [14]. When it comes to detecting fraud, methods like logistic regression, decision trees, and neural networks are invaluable. Unsupervised learning methods, such as clustering and anomaly detection, can detect data outliers that may suggest fraudulent behavior[15]. Methods like k-means clustering and isolation forests can be used to find anomalies in transaction data. One subfield of machine learning called "deep learning" relies on multi-layered neural networks[16]. The most effective methods for capturing complicated patterns in large datasets are Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). These algorithms can handle large amounts of transaction data and can detect complex fraud schemes.

III. METHODOLOGY

One of the most rapidly expanding markets is the private insurance industry. The last ten years have seen tremendous changes driven by this fast expansion. Jewellery, cars, health/life, and houses are just a few examples of the high-value assets that can be insured these days. In order to maximize profit while fulfilling their clients' claims, insurance companies are leading the way in implementing cutting-edge operations, processes, and mathematical models. Conventional approaches that rely solely on human-in-the-loop models are laborious and prone to errors.

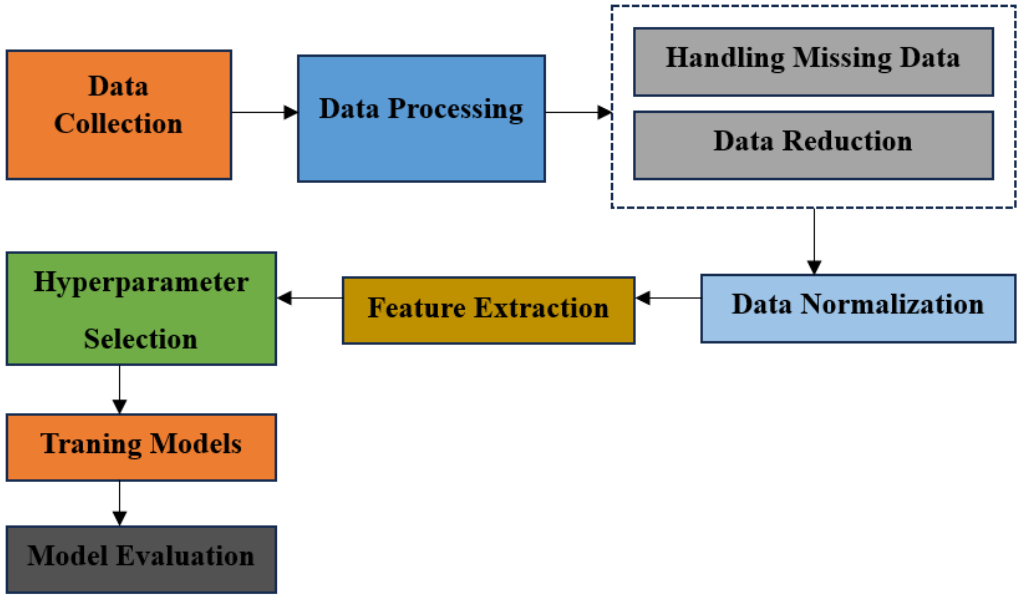

Fig. 1. Process Flow Diagram for this study

We explore the complexities of data collecting at each stage of the model evaluation process in the sections that follow. The process flow for the suggested models is illustrated in Figure 1 with great detail.

A. *Risk Management and Fraud Detection:*

Financial services have greatly improved proactive risk management and fraud prevention tactics with the transition towards AI and predictive analytics. The majority of the time, when it came to risk assessment and fraud detection, the conventional wisdom was that problems could only be addressed after the fact. It may be difficult for these systems to properly detect and counteract new threats if they are dependent on static rules and past data. On the other hand, systems driven by AI are great at taking the initiative. These systems are able to examine data in real-time and detect any dangers before they happen because they use sophisticated predictive analytics. When trained on large datasets, machine learning models can spot irregularities or red flags that could mean fraud or new financial dangers are on the horizon[17]. For

example, predictive models can accurately assess external causes, changes in behaviour, and patterns of transactions to foretell possible risks. Rather than responding to problems as they emerge, financial institutions can now take proactive steps and modify their strategy accordingly. Personalization of security protocols is another example of the proactive character of AI-driven risk management. With the use of AI, businesses may take a more nuanced approach to client and transaction-specific risk assessments and fraud protection strategies. False positives, which occur frequently in less tailored systems, can be lessened through this personalization, which aids in addressing unique risk profiles. Institutions may develop more precise and efficient risk mitigation measures if they can foresee possible dangers and adjust in real-time. In addition, risk management methods can be continuously monitored and adjusted with the use of AI and predictive analytics. Artificial intelligence systems can adapt on the fly, unlike more conventional approaches that may necessitate upgrades and human involvement on a more regular basis. With this capacity for constant monitoring, new vulnerabilities and threats can be addressed instantly, keeping the system more secure and resilient. In general, financial services that use AI and predictive analytics promote a preventative stance towards risk management and fraud. A more secure and efficient operational environment can be achieved by financial institutions by using advanced algorithms and real-time data analysis to predict and handle possible issues before they worsen.

B. Model Training

1) LR:

One or more input variables (multiple LR) can be utilized to model the mathematical relationship between the output variable and linear regression (LR). A linear relationship between the input characteristics and the response variable is assumed in LR. Equation 1 can be used to represent LR:

$$x = n_0 + n_1 y_1 + n_2 y_2 + \cdots + n_b y_b \quad (1)$$

as $[y_1, y_2, \dots, y_b]$ denotes the input features and $[n_1, n_2, \dots, n_b]$ denotes the regression coefficients, with y representing the target variable and bo the y-intercept. Finding the coefficients is a common application of the gradient descent method, which involves iteratively minimizing the sum of the squared errors from a set of randomly chosen coefficient values. A regression problem is the ideal application for LR, as the name indicates. When the dataset can be easily separated into linear components and the algorithm is straightforward to apply, LR become extremely valuable. One of the biggest problems with training ML algorithms is overfitting, which happens when a model does very well during training but can't generalize its predictions because it relies too much on specific input features. Itching and outliers are two factors that can affect LR's performance.

2) Decision Tree:

To solve classification and regression issues, a decision tree (DT) might be employed. A decision tree (DT) is like an ow chart in that it uses split points from the input features to break down complex judgments into a couple of simpler ones. The node at which decisions are made is known as a decision node. Nodes that do not undergo any additional splitting are referred to as leaf nodes. For regression issues, the prediction is based on the mean of all the elements in the leaf node. The projected sets of classes are shown by the leaf nodes in classification problems. Simple visual aids, like a tree diagram, make it easy to explain DTs and shed light on how they arrive at predictions. The problem is that it's easy for a single DT to over- or under-estimate a topic.

3) SVM:

The majority of support vector machines (SVMs) use their classification capabilities for classification problems, but they can also perform regression, albeit under the name support vector regression (SVR). In order to maximize margins between classes, support vector machines (SVMs) use hyperplanes for class division[18]. It is possible to use linear, polynomial, or radial basis function (RBF) kernels to linearly separate the inputs in high-dimensional feature spaces. The lengthy training time is a major issue with SVM. Big datasets might not be a good fit for SVM.

4) KNN:

In spite of its flexibility, k-nearest neighbor (KNN) is typically reserved for classification problems instead of regression. Bypassing a dedicated training phase is a hallmark of KNN, which is a kind of lazy learning. Improved prediction accuracy is achieved by identifying the k closest neighbors of a newly added data point using a distance

measure, often the Euclidean distance. Subsequently, it will be distributed to the group comprising the majority of the adjacent residents. This method, which is also called a 3-NN algorithm, is shown here with the value of k set to 3. For this example, the new item in green's three nearest neighbors consists of two things from the orange class and one object from the blue class. So, the new green item will be given to the orange class.

5) RF:

The aggregation of numerous decision trees is what allows random forest (RF) to make predictions. Here, the trees are constructed using a variety of bootstrap samples, also known as samples with replacements, and the bagging approach is employed. In regression, the average value of the predictions from all the trees is used for aggregation, while in classification, the majority vote across the trees is used. As an example of ensemble ML, RF takes a number of different ML models and uses them together to create a single model with better predicted performance. The thinking behind this method is quite similar to polling a group of experts for their thoughts on a topic and then using their votes to settle on a course of action. In a similar vein, XGboost, or a gradient boosting technique, makes use of numerous DTs; the main distinction is that, while building each tree, gradient boosting takes into account the mistakes made by the trees before it. Both methods significantly lessen the likelihood of overfitting when contrasted with the basic DT model.

IV. Results and Discussion

The primary goal of this proposed is to provide a thorough and all-encompassing analysis of AI methods employed for detecting financial fraud. These papers will later undergo further scrutiny. The findings of this study demonstrate that AI-based approaches are capable of detecting financial fraud. Specifically, the AI technique makes a huge leap forward in this area by greatly improving the efficiency and accuracy of fraud pattern detection.

TABLE I. Performance Prediction(%)

Metric	Accuracy	Precision	Recall
Linear Regression	0.8659	0.8415	0.8542
Decision Tree	0.8946	0.8728	0.8860
Support Vector Machine	0.9167	0.8947	0.9023
K-Nearest Neighbor	0.8837	0.8624	0.8739
Random Forest	0.9023	0.8805	0.8911

These five machine learning models are presented in table 1 together with their respective performance metrics: Random Forest, Linear Regression, Decision Tree, and Support Vector Machine (SVM). Each model's three most important performance measures are displayed in the table: Precision, Recall, and Accurac.

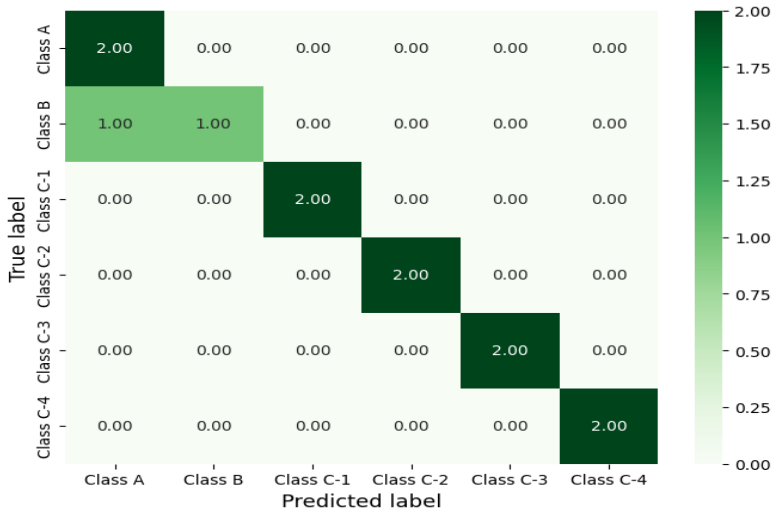

Fig. 2. Confusion Matrix for Predicting Fraud Detection in Banking

A confusion matrix summarizes the results of a classification model's predictions, as shown in Figure 2. When dealing with an imbalanced dataset or many classes, as is often the case in fraud detection, it is helpful for evaluating a model's accuracy. There are some misclassifications among various sorts of transactions, however the model does a good job for certain groups (probably high-risk fraud situations). This research can help banks improve their models, which in turn reduces financial risks and boosts consumer satisfaction.

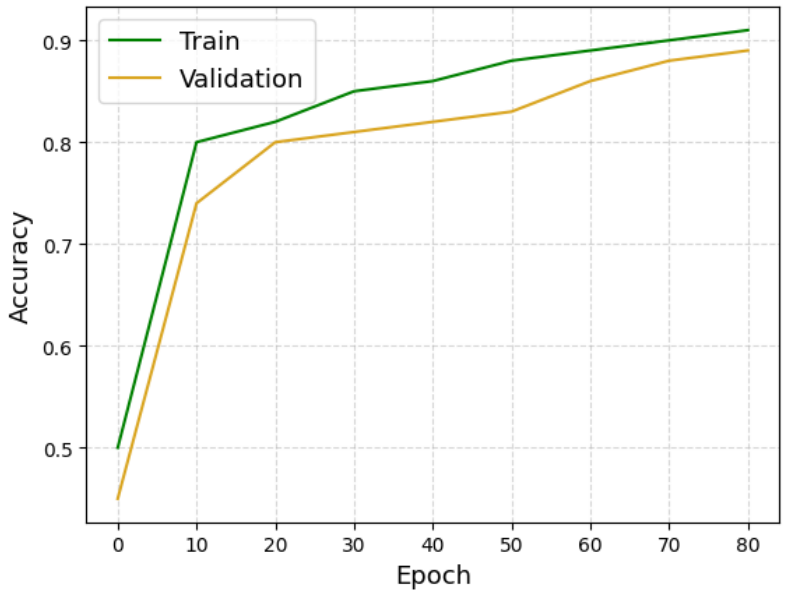

Fig. 3. Training and Validation Accuracy for SVM Model Predicting Fraud Detection in Banking

A banking-related fraud detection model with excellent training and validation accuracy is shown in figure 3. It seems like the model is ready for deployment because it learns and generalizes nicely. But to make sure the model keeps detecting fraud correctly over time, additional performance measures and continuous monitoring are needed.

Fig. 4. Training and Validation Loss for SVM Model Predicting Fraud Detection in Banking

An SVM model that predicts the detection of banking fraud suffers a loss during training and validation, as shown in Figure 4. In this case, the validation loss is 0.38 and the training loss is 0.27.

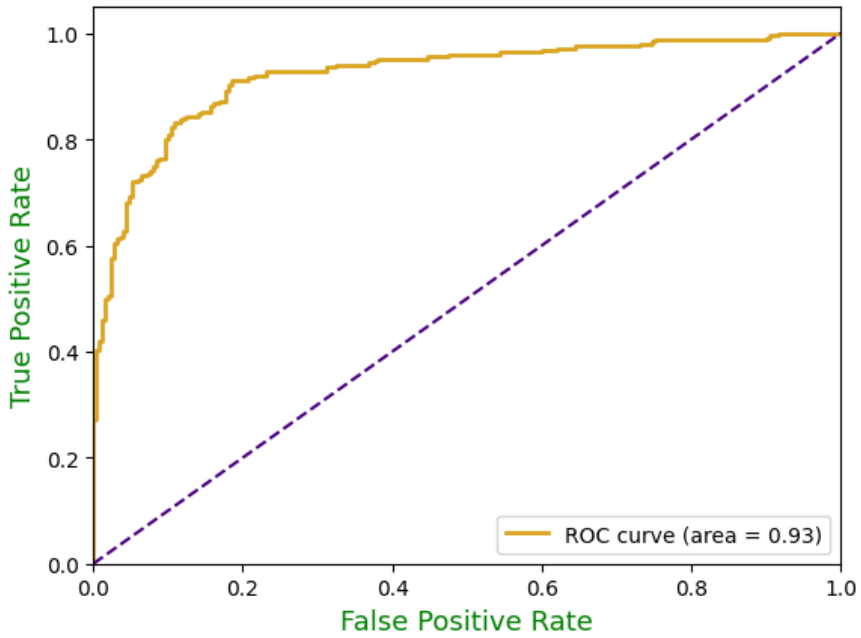

Fig. 5. ROC Curve for the Proposed Model

Bank fraud detection models use receiver operating characteristic curves, as seen in Figure 5. An area under the curve (AUC) of 0.93 shows that the model is very good at differentiating between real and fraudulent transactions; this is indicated by the yellow line that depicts the model's performance. The x-axis displays the number of false positives, and the y-axis shows the number of true positives. The performance of the model is improved when the AUC is greater, closer to 1. A random classifier without predictive power (AUC = 0.5) is represented by the dashed line.

V. Conclusion and Future Directions

Using AI-powered solutions to combat fraud and enhance risk assessment, financial services institutions are drastically altering their asset protection strategies while remaining compliant with regulations. As AI-enabled technologies become more commonplace, they will progressively supersede risk management and fraud detection methods that rely on static models and historical data. Machine learning algorithms are used by such state-of-the-art systems to scan through massive amounts of data for patterns and anomalies that may indicate fraud or emerging threats. Machine learning-based fraud detection systems can improve their performance over time by absorbing more and more data. By continuously tracking transactional behavior, customer relationships, and market patterns, these systems can uncover questionable activity that might otherwise go unnoticed by conventional methods. Financial institutions may improve fraud detection and decrease false positives with this preventative strategy, which also reduces operational stress. Artificial intelligence allows for more comprehensive and dynamic analysis in risk assessment. For a more comprehensive view of potential risks, machine learning models can integrate a wide variety of data sources, such as financial records, external factors (such as economic indicators), and social media activity. By taking this proactive measure, institutions can strengthen their stability and resilience by mitigating risks before they even occur. Using support vector machines (SVMs), we trained a model to anticipate financial fraud detection with a maximum accuracy of 91.67 percent.

REFERENCES

[1] R. Rambola, P. Varshney, and P. Vishwakarma, “Data mining techniques for fraud detection in banking sector,” *2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018*, pp. 1–5, 2018, doi: 10.1109/CCAA.2018.8777535.

[2] N. Pol and S. Agarwal, “Online Transaction Fraud Detection : Exploring the Hybrid SSA-TCN-BiGRU Approach,” *2024 2nd World Conf. Commun. & Comput.*, pp. 1–6, 2024, doi:

10.1109/WCONF61366.2024.10692254.

[3] T. Porkodi, "An Automatic ATM Card Fraud Detection Using Advanced Security Model Based on AOA-CNN-XGBoost Approach," *2024 Int. Conf. Electron. Comput. Commun. Control Technol.*, pp. 1–7, doi: 10.1109/ICECCC61767.2024.10593851.

[4] K. Yamini, V. Anitha, S. Polepaka, R. Chauhan, Y. Varshney, and M. Singh, "An Intelligent Method for Credit Card Fraud Detection using Improved CNN and Extreme Learning Machine," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jun. 2023, pp. 810–815. doi: 10.1109/ICCES57224.2023.10192774.

[5] N. Soltani Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," *Appl. Soft Comput. J.*, vol. 24, pp. 40–49, 2014, doi: 10.1016/j.asoc.2014.06.042.

[6] N. M. Reddy, K. A. Sharada, D. Pilli, R. N. Paranthaman, K. S. Reddy, and A. Chauhan, "CNN-Bidirectional LSTM based Approach for Financial Fraud Detection and Prevention System," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, Jun. 2023, pp. 541–546. doi: 10.1109/ICSCSS57650.2023.10169800.

[7] R. Rajkumar, N. Kogila, S. Rajesh, and A. R. Begum, "Intelligent System for Fraud Detection in Online Banking using Improved Particle Swarm Optimization and Support Vector Machine," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jun. 2023, pp. 644–649. doi: 10.1109/ICCES57224.2023.10192690.

[8] X. Kewei, B. Peng, Y. Jiang, and T. Lu, "A Hybrid Deep Learning Model for Online Fraud Detection," *2021 IEEE Int. Conf. Consum. Electron. Comput. Eng. ICCECE 2021*, no. Iccece, pp. 431–434, 2021, doi: 10.1109/ICCECE51280.2021.9342110.

[9] A. Ruchay, E. Feldman, D. Cherbadzhi, and A. Sokolov, "The Imbalanced Classification of Fraudulent Bank Transactions Using Machine Learning," *Mathematics*, vol. 11, no. 13, pp. 1–15, 2023, doi: 10.3390/math11132862.

[10] K. Rathor, S. Vidya, M. Jeeva, M. Karthivel, S. N. Ghate, and V. Malathy, "Intelligent System for ATM Fraud Detection System using C-LSTM Approach," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, IEEE, Jul. 2023, pp. 1439–1444. doi: 10.1109/ICESC57686.2023.10193398.

[11] S. Kakkar, "Analysis of Discovering Fraud in Master Card based on Bidirectional GRU and CNN based Model," *2023 Int. Conf. Self Sustain. Artif. Intell. Syst.*, no. Icssas, pp. 50–55, 2023, doi: 10.1109/ICSSAS57918.2023.10331770.

[12] K. Prabhakar, M. S. Giridhar, A. Tatia, T. M. Joshi, S. Pal, and U. S. Aswal, "Comparative Evaluation of Fraud Detection in Online Payments Using CNN-BiGRU-A Approach," *Int. Conf. Self Sustain. Artif. Intell. Syst. ICSSAS 2023 - Proc.*, no. Icssas, pp. 105–110, 2023, doi: 10.1109/ICSSAS57918.2023.10331745.

[13] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," *arXiv Prepr. arXiv1009.6119*, 2010, doi: 10.1016/j.chb.2012.01.002.

[14] M. Â. L. Moreira *et al.*, "Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems," *Procedia Comput. Sci.*, vol. 214, no. C, pp. 117–124, 2022, doi: 10.1016/j.procs.2022.11.156.

[15] *et al.*, "Ai-Powered Fraud Detection in Banking: Safeguarding Financial Transactions," *Am. J. Manag. Econ. Innov.*, vol. 6, no. 6, pp. 8–22, 2024, doi: 10.37547/tajmei/volume06issue06-02.

[16] V. Sambrow, K. I.-E. R. of S. and, and undefined 2022, "Integrating Artificial Intelligence in Banking Fraud Prevention: A Focus on Deep Learning and Data Analytics," *Stud. Sambrow, K IqbalEigenpub Rev. Sci. Technol. 2022•studies.eigenpub.com*, vol. 6, no. 1, p. 2022, 2022, [Online]. Available: https://studies.eigenpub.com/index.php/erst/article/view/42

[17] A. Yuille, "AI-Powered Financial Services : Enhancing Fraud Detection and Risk Assessment AI-Powered Financial Services : Enhancing Fraud Detection and Risk Assessment with Predictive Analytics," no. August, 2024, doi: 10.13140/RG.2.2.23580.09603.

[18] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, "Machine learning approaches for EV charging behavior: A review," *IEEE Access*, vol. 8, pp. 168980–168993, 2020, doi: 10.1109/ACCESS.2020.3023388.