Effective Analysis and Visualization of Fraud Detection Patterns Through Data Mining and Classification using MLP and Hybrid Deep Learning Model

1st **Jagdesh Tulsi** *

2nd **K. Mansoor Ali***

Abstract— **Fraudulent transactions represent serious risks in a range of industries, needing enhanced detection and classification methodologies. The investigation goes into the issue of unauthorized transaction classification, emphasizing the necessity of accurate identification to prevent monetary damage as well as preserve system integrity. Utilizing fraud historical data, the study employs powerful data-mining and visualization approaches to discover hidden trends that accompany fraudulent behavior. For classification, two reliable deep learning models, Multilayer Perceptron (MLP) and a hybrid model combining 1D Convolutional Neural Network (1D-CNN) and Long Short-Term Memory (LSTM), have been utilized. The dataset under review exhibits a built-in imbalance, with an uneven proportion of legal and wrongful transactions. The Synthetic Minority Over-Sampling Technique (SMOTE) is implemented to address the asymmetry and eliminate biases in the training of models. This leads to an even more solid and unbiased training and evaluation dataset. Furthermore, the development of two distinct deep-learning models for classifying illicit transactions is at the core of this work. The Multilayer Perceptron (MLP) performs extremely well, with an accuracy of 99.64%, precision of 99.47%, recall of 99.63%, and F1-score of 99.64%. The 1D-CNN-LSTM hybrid model, on the other hand, has a more effective in accuracy of 99.76%, precision of 98.94%, recall of 99.57%, and an F1-score of 99.46%. Therefore, the comparative evaluation of these models suggests their effectiveness in detecting fraudulent transactions, although a better knowledge of their performance is essential. While the MLP model excels in precision and recall, the 1D-CNN-LSTM hybrid model improves accuracy and recall. The selection between these classifications is decided by the application's particular requirements. If accuracy and recall are crucial, the MLP model might be employed. Yet, if an appropriate ratio of accuracy and recall is required, the 1D-CNN-LSTM hybrid model arises as a promising rival. In addition, this work utilizes unique analytical visualizations together with deep learning models to address the issue of bogus transaction classification. The findings reveal the usefulness of both the MLP and the 1D-CNN-LSTM frameworks, enabling a more thorough knowledge of their strengths and shortcomings. The findings enable those who work to make sensible decisions depending on their application's unique aims, enabling ongoing attempts to enhance systems for identifying fraudulent activity.**

Keywords— MLP, Transactional Fraud Detection, LSTM, SMOTE, 1D-CNN-LSTM

I. INTRODUCTION

The complicated nature of fraudulent conduct and technical advancements have prompted an evolution in the detection of credit card fraud techniques. At first, rule-based systems were mostly used for detection. These systems would identify transactions based on predetermined criteria, such as unusual expenditure habits or transactions from unidentified sources. However, the flexibility of these systems was limited to continue keeping up with the ever-evolving techniques employed by scammers [1]. A new wave of innovation in identifying fraud was brought about in the late 20th century by the development of computational models and data analysis. The research effort used statistical techniques for examining transaction data and detecting anomalies that may point to fraud. Machine learning techniques became more and more common in identifying credit card fraud as the digital age progressed. Researchers began using statistical algorithms to analyze transaction data and find anomalies that may point to fraud. As the digital era progressed, machine learning techniques gained importance in the discipline of fraud with credit card identification. Machine learning techniques—a branch of artificial intelligence—became well-liked for their ability to identify minute patterns and irregularities in transaction data. Many noteworthy studies, such as those on learning systems and neural networks, highlight the trend toward increasingly sophisticated modeling techniques [2]. The use of models based on deep learning and other complex algorithms in conjunction with big data analytics has made fraud detection more accurate in recent years. Overall, the lengthy history of discovering methods for detecting fraud involving credit cards demonstrates a never-ending search for ever-more-adaptable and efficient solutions, making use of developments in processing speed and analysis to stay ahead of the ever-changing landscape of fraudulent activities. To obtain the maximum possible accuracy, this article aims to examine current methodologies and their efficacy and to provide the best methods for preprocessing and complex model application to existing datasets [3].

II. LITERATURE REVIEW

Recent studies [1] focus on minimizing false alarms" and address the escalating popularity of e-wallets, emphasizing their role in facilitating a range of financial activities while eliminating the need for physical currency or cards. Acknowledging the vulnerabilities of user financial information to threats like phishing, malware, and social engineering, the study highlights the proactive measures taken by fintech platforms through intelligent fraud detection mechanisms. The research focuses on leveraging cutting-edge machine learning techniques to detect fraudulent activity, utilizing data from Turkey's leading e-wallet platform. Notably, the LightGBM approach emerges as the most effective, achieving a 97% detection accuracy, and successfully reducing false alarms from 13,024 to 6,249. The paper

* Authors contributed equally to this work.

contributes by presenting a machine-learning model tailored for smaller fraud detection teams, emphasizing practicality and efficiency in mitigating e-wallet transaction fraud. Integrating Machine Learning Algorithms with Quantum Annealing Solvers for Online Fraud Detection [2] addresses the evolving landscape of fraud detection by proposing a novel framework that integrates quantum machine learning (QML) with traditional machine learning approaches. The traditional systems often detect fraudulent activities post-occurrence, lacking real-time efficacy. Given the highly imbalanced nature of fraud data, the study leverages quantum annealing solvers to enhance Support Vector Machine (SVM) performance. Evaluating the framework across two datasets—a moderately imbalanced Israel credit card transactions set and a highly imbalanced bank loan dataset—the results showcase the quantum-enhanced SVM's superior speed and accuracy in the context of time series data. Feature selection is identified as a crucial factor, significantly enhancing detection speed while maintaining a marginal impact on accuracy. The study underscores the potential of QML applications for highly imbalanced, time-series data and highlights the nuanced trade-offs between speed, accuracy, and cost when choosing an approach for different dataset types. Another important research [3] on Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms targets the pressing issue of credit card fraud in the context of increasing online transactions. The literature review acknowledges the rising misuse of credit cards and the consequential financial losses incurred by both cardholders and financial institutions. While existing literature explores various machine learning approaches such as Extreme Learning Methods, Decision Trees, Random Forest, Support Vector Machine, Logistic Regression, and XG Boost, the paper highlights the persistent challenge of low accuracy and emphasizes the need for state-of-the-art deep learning algorithms to reduce fraud losses effectively. The study focuses on the recent developments in deep learning and conducts a comparative analysis between traditional machine learning and advanced deep learning algorithms using the European card benchmark dataset. The empirical analysis reveals significant improvements in accuracy, f1-score, precision, and AUC curves, with the proposed model outperforming both machine learning and deep learning algorithms in credit card fraud detection, providing a valuable contribution to the field. The investigation [4] draws attention to the urgent problem of fake news's explosive growth, which is defined as the deliberate dissemination of material that may be proven to be false for fraudulent ends. To create an efficient false news identification system, the research investigates unique techniques built on both Machine Learning (ML) and Deep Learning (DL), recognizing disinformation's serious social impact on political division and public confidence. The main goal is to determine which model performs best in terms of accuracy, and this leads to the suggestion of the "Optimized Convolutional Neural Network model (OPCNN- FAKE)". It thoroughly compares and contrasts DL models like RNN and LSTM with classic ML approaches like Decision Tree, Logistic Regression, K Nearest Neighbor, Random Forest, SVM, and Naive Bayes. The research's resilience is enhanced by the use of methods for optimization, grid search, hyperopic, and feature extraction techniques like TF-IDF and N-gram. When findings are validated using the evaluation criteria of accuracy, precision, recall, and F1-measure, it is evident that OPCNN-false performs significantly better than other models for the identification of false news over a variety of data.

In yet more fascinating work [5], the researchers addressed the difficult problem of Shill Bidding (SB) in e-auctions, where fraudulent bidders are utilized by vendors to artificially raise ultimate prices. They also considered Real-Time Shill Bidding Fraud Identification Empowered by Fussed Machine Learning. This form of fraud is particularly elusive due to its mimicking of normal bidding behavior. The paper proposes a novel fusion-based model to enhance real-time detection of shill bidding, aiming to protect bidders from financial losses and curb the benefits sellers gain from this fraudulent practice. The model is structured into training and validation phases, with three sub-modules. Two machine learning algorithms, Support Vector Machine (SVM) and Artificial Neural Network (ANN), operate in parallel to predict bidding fraud. The predictions from SVM and ANN serve as inputs to a fuzzy-based fusion module, which then determines the actual output. The proposed model demonstrates a high prediction accuracy of 99.63%, surpassing state-of-the-art methods in the literature. By integrating machine learning algorithms and fuzzy logic, this research contributes a robust solution for real-time shill bidding fraud detection in e-Auctions. In [6] titled as "Review of Loan Fraud Detection Process in the Banking Sector Using Data Mining Techniques" the author critically addresses the escalating issue of fraud in the banking industry amid the era of digital transformation. The pervasive nature of financial fraud, with its substantial economic costs, underscores the urgent need for effective strategies and methods to prevent and detect such activities. The paper emphasizes the role of modern technology, innovation, and global communications in both the proliferation of fraud and the development of deterrent technologies to counteract it. Recognizing the significance of data mining techniques, the paper explores their widespread use in the prevention and detection of financial fraud. It highlights the alignment of fraud detection methodologies with data mining norms, encompassing crucial steps such as feature selection, representation, data gathering and management, pre-processing, comment, and summative evaluation. The literature review within the paper provides a comprehensive comparison of various fraud detection strategies, focusing specifically on loan banking and financial fraud. Recent research [7] recommends leveraging the related-party transactional knowledge graph to detect fraudulent financial transactions. It addresses the continuous difficulty of detecting financial fraud, with a particular emphasis on the often hidden and manipulable domain of transactions between related parties (RPTs) among corporations. The abstract emphasizes the limits of typical quantitative research approaches that regard each corporation as a separate entity, ignoring the complicated linkages and transactions between connected parties. The study describes the novel usage of an information graph to convey and mine significant hidden information from massive, linked data, providing a fresh viewpoint on how knowledge is represented in identifying fraud. The project intends to use the RPT knowledge tree to improve fraud in finance detection by integrating characteristics related to transaction magnitude and category. The study carried out on Chinese listed businesses from 2000 to 2019 shows that these qualities increase the financial fraud identification effectiveness, implying that elements such as the kind, magnitude, and regularity of RPTs can be used as indications of fraud. The identified feature importance emphasizes the significance of regulatory attention to specific aspects, particularly loan-based RPTs and the total number of RPTs, offering valuable insights for practitioners and regulators

in the field of financial fraud detection. Also, the study [8] based on The Combining Control Rules, Machine Learning Models, and Community Detection Algorithms for Effective Fraud Detection" addresses the escalating concern of fraud in financial institutions, emphasizing the significant financial losses and growing risks faced by customers. Acknowledging the need for effective fraud management, the paper explores the integration of data analytics capabilities to identify potential fraud in real-time. The abstract highlights the use of supervised machine learning models and graph/network analysis approaches as key strategies for fraud detection. The paper presents the findings of two studies, utilizing both open-source data and information from an anonymous financial institution. The results emphasize the synergistic benefits of combining machine learning models with control rules, showcasing improved accuracy in fraud detection. Additionally, the research underscores the pivotal role of graph analysis techniques, particularly community detection, in uncovering the intricate network of fraudsters. This literature review within the paper contributes valuable insights into the effectiveness of combining diverse approaches for a comprehensive and enhanced fraud detection system in the financial domain. In a study [9] focuses on Real-time credit card fraud detection using Streaming Analytics researchers delve into the critical issue of credit card fraud in the context of the booming internet-driven e-commerce landscape. The abstract outlines the convenience and benefits of credit card usage but highlights the security challenges posed by theft, loss, or misuse, adversely affecting cardholders, banks, and merchants. The paper introduces streaming analytics as a solution for real-time credit card fraud detection, emphasizing its capability for time-based processing of data to enable near real-time decision-making. Instead of isolating specific transactions, the proposed solution utilizes historical transaction data to model and detect fraudulent patterns, providing a more comprehensive and proactive approach to fraud prevention. The literature review within the paper likely explores existing methods and technologies in credit card fraud detection, positioning streaming analytics as an innovative and efficient solution for addressing the evolving challenges in the realm of online and offline payments. In [10] the paper mentions a Novel Approach to Credit Card Fraud Detection System Using Machine Learning Techniques. It addresses the escalating issue of credit card fraud, particularly with the increasing reliance on online purchases and the surge in fraudulent activities amid the COVID-19 pandemic. The abstract underscores the significance of finding effective methods for detecting scams in online systems, highlighting the challenges posed by the use of credit card credentials in contactless transactions. The paper acknowledges the need for credit card companies to identify fraudulent transactions promptly to protect customers from unauthorized charges. The abstract suggests that several theories exist for detecting these frauds, and the study aims to contribute a novel approach using semi-supervised machine learning models. It emphasizes the importance of dealing with imbalanced datasets, a common challenge in fraud detection. The literature review within the paper likely explores existing theories and methods for credit card fraud detection, positioning the proposed semi-supervised machine learning approach as a novel and effective solution to address the evolving landscape of online fraud in credit card transactions. Credit card fraud events take place frequently and then result in huge financial losses. The number of online transactions has grown in large quantities and online credit card transactions hold a huge share of these transactions. Therefore, banks and financial institutions offer credit card fraud detection applications much value and demand. Fraudulent transactions can occur in various ways and can be put into different categories. Another research [11] focuses on four main fraud occasions in real-world transactions. Several algorithms based on machine learning are used to tackle each scam, and an assessment is used to determine which approach works best. This assessment offers thorough guidance for choosing the best algorithm for the kind of fraud and supporting the assessment with a suitable performance metric. Real-time credit card fraud detection has been tackled as a significant area of focus in this project. To determine if a certain transaction is legitimate or fraudulent, predictive analytics is used in conjunction with an API module and machine learning models that have been put into place. It also evaluates a new approach that successfully deals with skewed data distribution. A confidentiality disclosure agreement states that the financial institution provided the data utilized in the tests. In the rapidly evolving landscape of online banking, the prevalence of fraud poses a significant challenge to the security of financial transactions. The field of identifying and avoiding fraud has seen a significant transformation with the incorporation of machine learning and big data analytics methodologies. Studies have acknowledged the growing significance of efficiently evaluating large amounts of transactional information to reveal concealed patterns and market trends. The application of principle component analysis (PCA) for extracting and reducing features has become a prominent practice, enabling the effective management of large datasets. Moreover, the utilization of self-organizing maps in the process of generating and detecting user patterns has become increasingly prominent as an advanced method to detect fraudulent behaviors. Prior research emphasizes the necessity of creative approaches that combine these technologies to improve the safety features of Internet banking and safeguard users from unwanted access and fraudulent activities [12]. The existing body of research emphasizes the significant and transformational influence of big data analytics (BDA) on the banking sector. The banking industry, along with other industries, has acknowledged the necessity of utilizing big data analytics to gain a strategic advantage due to the rapid increase in population and the resulting large amounts of data. Multiple sectors, such as financial services, online commerce, and insurance companies, have progressively embraced advanced analytical methods to analyze and understand extensive datasets, resulting in more knowledgeable decision-making procedures. Big data analytics is essential for detecting fraud, especially in the field of online banking where the increase in electronic transactions has raised the danger of fraudulent operations. An essential aspect highlighted in the literature is the significance of proficient examination of client data for banking institutions to acquire practical knowledge, encompassing comprehension of financial transactions as well as detection and reduction of risks linked to calamities. Moreover, the implementation of analytics for big data in the banking industry is considered an approach to attaining advertising objectives and staying relevant in the ever-evolving field. As the banking sector in India adopts the application of big data analytics, it is projected to not only boost decision-making but also secure a long-term place in the market amidst increased competition [13]. The studies related to the contextual analysis of artificial intelligence (AI) algorithms, especially Quadratic-Discriminant-Analysis (QDA), Logistic-Regression (LR), and Support Vector Machines (SVM), in the

field of identifying credit card fraud, highlight the vital requirement for robust mathematical models for stopping fraudulent transactions. With the rising incidence of unauthorized purchases made online, credit card issuers confront the difficulty of effectively recognizing and combating fraudulent activity to defend both users and their financial well-being. Previous investigations has looked into the implementation of different methods of machine learning to achieve optimum fraud detection, highlighting the role of modeling and data preparation in boosting accuracy. QDA, LR, and SVM have been recognized as important candidates in this arena, each having qualities and drawbacks. Investigators have examined the usefulness of these algorithms, analyzing how they perform based on varied datasets and the creation of feature extraction methods. The research underlines the significance of a complete investigation to find the best suitable machine learning strategy for credit card fraud detection, including criteria such as accuracy of models, computational performance, and flexibility to emerging fraud trends [14]. The research done in the field of financial recognition of fraud underlines the rising relevance of Explainable-Artificial-Intelligence, or XAI, methodologies in assuring not only the accuracy of forecasting algorithms but also the integrity and clarity of their methods for making decisions. As banks and other financial institutions increasingly utilize complex machine-learning algorithms for identifying fraud, the requirement for focused user explanations becomes important. Prior studies have studied several XAI approaches, with Shapley values appearing as a suitable foundation for offering knowledge about model forecasts. The integration of ensemble forecasting algorithms with Shapley-based interpretations is a cutting-edge technique to find the right equilibrium between accuracy and understanding in the financial identification of fraud. The literature emphasizes the necessity of delivering global as well as local explanations to respond to the different demands of outside parties, allowing for a greater comprehension of individual predictions and uncovering the fundamental reasoning of the ensemble approach. This user-centered approach not only promotes trust in the prediction model but also allows interaction among data scientists, financial specialists, and users as a whole resulting in more successful fraud mitigation strategies [15]. In an additional study on online fraud discovering the issues caused by data disparities has been a focus point, with a specific emphasis on overcoming the dominance of authentic activities over fraudulent ones. The application of deep learning has come to be a strong method for successfully acquiring knowledge from unbalanced information, highlighting its promise in systems for detecting fraud. However, the detailed examination of thresholding strategies within the field of deep learning remained largely unexplored. Thresholding, as a way to update decision thresholds in learning models, provides an alternative to adjusting the distribution of data or the learning process itself. This study extends to the current research by carrying out a comparative evaluation of three thresholding techniques to get closer that utilize Receiver-Operating-Characteristic (ROC) Curve criteria, namely Nearest to (0,1) criterion, Youden Index (J), and max -G-Mean. The work focuses on analyzing the effectiveness of these strategies in a deep-learning-based fraud in online transaction detection system. The literature emphasizes the necessity of establishing an appropriate balance between enhancing fraud detection rates, especially True Positive Rate (TPR), and preserving the overall effectiveness of the model. The results of the study reveal that the nearest to (0,1) criterion beats the other two strategies, stressing the necessity of selecting the best decision-thresholding strategy for optimal outcomes in online systems for detecting fraud utilizing deep learning [16]. The special issue examines the most current breakthroughs in analytics for big data and the use of artificial intelligence (AI) in diverse industries, such as goods, markets for financial instruments, health, and environmental issues. It underlines data's exponential rise in the healthcare, higher education, and energy sectors. Combining AI with Big Data enhances prediction abilities, allowing for increased decision-making processes, policy creation, and management across domains while handling complex situations and generating important insights [17]. The article discusses the notion of large-scale data analytical service-orientated architecture (BASOA) as a strategy for leveraging big data analytics to boost the effectiveness of business intelligence (BI). It stresses the relevance of timing, predictability, and application in BI insight. The recommended plan fosters works on intelligence for business, large-scale data analysis, and BI, while also contributing to big data extraction and computing [18]. The study of literature covers current breakthroughs in analytics of big data, highlighting positive usage in healthcare, marketing, and the public sector while addressing concerns such as the accuracy of data and scalability. It also stresses new trends, technology, and potential clients, offering crucial knowledge for experts and instructors who work in this dynamic field [19]. It also stresses new trends, methodologies, and prospects for the future, giving vital information to scholars and instructors on this ever-changing topic [20]. This research study investigates the changing financial landscape, which is defined by digitalization and a necessity for fraud protection. It proposes a real-time architecture for recognizing transactional fraud that takes the use of analytics tools for big data including Spark, Kafka's, and h2o, in addition to the isolation forest unsupervised machine learning approach. Experiments with a large digital transaction dataset reveal the framework's stability, with a detection accuracy of 99% and an 87% precision rate, emphasizing its usefulness in identifying and combating fraud in a perpetually changing financial environment [21]. Because of the rapid growth of smartphones, the Internet, and financial technology, this study focuses on recognizing fraudulent activities from the perspective of internet-based purchases. It provides a breakthrough approach based on artificial intelligence and massively parallel analytics algorithms to spot fraud in the large volumes of data produced by e-commerce. The efficiency, scalability, and accuracy, of the approach for real-time processing are proved by experimental data from a Chinese e-commerce firm [22]. This focuses on the major problem of credit card fraud in online retail, with an emphasis on the creation and deployment of a system for fraud detection in a big e-tail merchant. It addresses the usage of human and automated categorization methods, offers insights into system creation, and evaluates various machine-learning approaches. Provides useful information for researchers and professionals wanting to develop data mining-based fraud identification systems, thereby improving fraud detection efficiency and overall effectiveness [23]. This study tackles the problem of Internet banking detecting fraud by offering a tailored alert model that utilizes sequence patterns extracted from every user's transaction log. The approach tries to improve the efficacy of identifying fraud by customizing it to specific users and detecting deviations from their regular transaction patterns. It sequences transaction records and raises a notification if inbound transactions stray considerably from the preset patterns.

Experiment findings reveal that our customized model outperforms traditional based on rules and Markov chain-based models, indicating that it has real-world implications in detecting online banking fraud [24]. This study tackles the widespread problem of fraud among consumers in online buying, as well as a lack of data-driven information concerning warning signs. Data mining techniques, especially decision trees, are used to uncover trends and danger signals in transaction information from a large online store. The findings provide important insights for improving fraud protection tactics and techniques across real-world and theoretical contexts, contributing to the ongoing fight against e-commerce fraud [25]. This study dives into the methodologies used by financial institutions and credit card firms for recognizing fraud, to identify the major characteristics linked with fraudulent actions. Using large datasets, the study uses the correlation graph in a descriptive analysis to identify features with the strongest link to fraud. Machine learning methods such as decision tree models, Linear Regression (LR), and Logistic Regression, among others, are then used to detect fraud with remarkable accuracy. This study improves the personnel working with fraudulent-related issues' comprehension of fraud detection [26]. The issues of legitimacy and security in online tests, showing differences in moderator location and authorization from users. It investigates several authentication methods such as unimodal, multimodal, hardware interaction, and data visualization. The study offers fraudulent activity on Online Tests (FDOT) and Habit Identification via visualization techniques (BIVT) to improve the security and efficacy of online exams, to reduce malpractice and disobedience [27]. ATOVis is a tool for visualization meant to improve detection of fraud, especially in the context of Account Takeover (ATO) trends in the Finance area, particularly online commerce. ATOVis is a powerful tool for visually evaluating and identifying suspicious patterns in data related to transactions. For ATO identification, the tool offers a task level of abstraction, two visualization models, and multiscale timelines to provide an information overview. User review with detection of fraud specialists proves its usefulness in enhancing analysts' capacity to spot certain fraud behaviors [28]. The examined study underlines the vital role of visualizing of data in identifying fraud by illustrating how it aids in early dataset reviews. The study explains how visualization may expose probable mistakes and abnormalities, increasing data preprocessing and modeling for improved and more efficient fraud detection in transactions involving mobile payments employing the PAYSIM database as an example. It underscores the necessity for representations of data in every identified fraud endeavor to verify data validity and relevance before the full study [29]. It examines important algorithms for classification, such as Naive Bayes, Decision Tree, and Random Forest, in anticipation of and recognition of victims in Indonesia based on socio-demographic information. The findings demonstrate that Naive Bayes and Decision Tree topped the Random Forest model, with an accuracy of 77.3% against 76.8%, underscoring the necessity of using correct techniques for classification within the framework of artificial intelligence (AI) for fraud detection [30]. The research presented here underlines the expanding threat of financial deceit and the requirement for creative techniques to resist developing fraud strategies. It offers the Naive Bayes approach (NB) as a previously discovered Machine Learning approach to fraudulent credit card detection (FCCD) and analyzes how it performs compared to that of existing ML techniques. The data reveal that NB beats similar classifiers when it comes to precision, recall, accuracy, f-measure, and low False Positive Rate (FPR), demonstrating its ability to serve as a key AI model for boosting the detection of fraud in the banking sector [31]. This paper tackles the topic of identifying fraud using debit cards in transactions over the internet and presents methods based on machine learning as well as deep learning for improved effectiveness. It compares the methods of (CNN), CNN with methods (GRU), and Adaptive Boosting (AdaBoost). The adoption of the Synthetic Minority Oversampling Technique (SMOTE) assists in the correction of dataset imbalances. In particular, the Convolutional Neural Network beats earlier published work when it comes to AUC-ROC, accuracy, recall, and precision providing a viable strategy for reliable credit card fraud detection [32]. The paper focuses on detecting e-commerce transactions that are fraudulently utilizing machine learning methods such as SVM, LR, Naive Bayes, Decision Trees, and Random Forests. The classifier utilizing Random Forests got the maximum accuracy at 99.94%, which makes it a feasible AI technique for identifying fraud and protecting customers' cash transactions [33]. It is focused on identifying fraud in online retailers and limiting the risk of financial loss by merging large-scale data mining (BDM) with information fusion technology (IFT). Its purpose is to establish an effective detection and prevention model (FDM) for B2B e-commerce enterprises by merging computer technology (CT), artificial intelligence (AI), and data mining (DM). The proposed IFT-based FDM surpasses known techniques such as (SVM) and (LRM) in identifying fraud with better accuracy. BDM technology is crucial in this approach, giving a possible tool for identifying fraud in electronic commerce and fostering the healthy expansion of e-commerce between businesses [34]. Fraud with credit cards and underlining the necessity for exact alerts to prevent losing cash. It contrasts a mixed ensemble of varied models for machine learning with methods based on deep learning and takes into account real-world restrictions such as finite investigators and the costs associated with alarming errors and undetected frauds. The paper tests such models using a huge transactional dataset from an important corporation in the Republic of Korea utilizing the champion-challenger architecture. The deep learning challenger overtook the champion in both offline and post-launch testing, demonstrating the possibility of models based on deep learning to improve identifying fraudulent activity accuracy [35]. This study focuses on increasing the identification of bogus transactions in payments made without cash, where the difficulty is to identify a few fraudulent incidents within the bulk of legitimate transactions. On the Fraud Train and Fraud Test datasets, deep learning and machine learning methods such as Local Outlier Factor (LOF) and Auto encoders were employed. LOF achieved 99.0% accuracy with few false positives, which makes it a viable strategy for properly categorizing forged transactions in the real-life environment [36]. This study describes a SpiHWO-based deep RNN fraud detection approach that employs the transformation of data, selecting features and a Deep RNN model. The technique established its efficiency in spotting fraud by reaching a high accuracy of 0.951 with higher sensitivity and specificity [37]. The research described here centers on countering illicit activity in payments by integrating nature-inspired hyperparameter change with supervised machine learning models for classification, using an upgraded XGBoost approach. The process includes instructing and confirming the models with a ten-fold cross-validation approach, and it also involves discovering unauthorized or counterfeit payments,

which aids in the ongoing endeavor to recognize and prevent fraudulent transactions across a wide range of territories, including e-commerce and medical facilities [38]. This paper tackles the challenge of identifying fraud using payment cards in imbalanced datasets [39][40] by offering a strategy for leveling the data set via the use of an improved conditional generative adversarial network model. The balanced data is then linked with a random forests technique to identify fraud more efficiently. The technique presents a realistic way to fight the rising occurrence of fraud by employing credit cards as get more popular [41]. This contribution tackles the challenge of identifying fraudulent usage of bank cards in unbalanced datasets where fraud data is substantially smaller than ordinary transactions. To tackle this issue, the research provides an improved oversampling technique that leverages a Variational Autoencoder Generative Adversarial Network (VAEGAN) to augment minority class knowledge. This methodology has more advantages than other oversampling approaches such as GAN, VAE, and SMOTE regarding Precision, F1_score, and numerous other indices. In recognizing credit card fraud, the new VAEGAN-based oversampling method was demonstrated to be a superior solution for dealing with imbalanced data [39]. The goal of this investigation is to explore the difficulty of recognizing huge transfers connected to telecom fraud. A GAN model is provided to calculate the chance of fraudulent activity for each big transfer, allowing institutions to take preventative steps. In trials, the model surpasses earlier classification techniques by employing a comprehensive eliminating auto-encoder and adversarial learning. Its real-world applications in two financial institutions have led to a loss reduction of roughly 10 million RMB over twelve weeks, greatly strengthening the company's image [42]. Using unsupervised machine learning, this study is concerned with spotting online payment frauds with shifting trends. For the detection of anomalies, a deep Auto-encoder and confined Boltzmann machine (RBM) approach is provided. The model, created using TensorFlow and H2O, analyzes regular transaction restoration by delivering measures such as the mean-squared-error, root-mean-squared-error, as well as area under the curve [43]. This piece of writing presents a method for identifying fraud involving credit cards in extremely skewed datasets utilizing a neural network-based classifier ensemble and hybrid data resampling. The ensemble classifier blends an artificial neural network using (LSTM) as well as (AdaBoost). Data balance is improved by using a hybrid resampling methodology that combines the synthetic minority oversampling technique with the edited nearest neighbor (SMOTE-ENN) method. The suggested LSTM ensemble surpasses existing methods in terms of both specificity and sensitivity, with values of 0.998 and 0.996, respectively. This method tries to detect unauthorized transactions with credit cards and reduce financial losses [40]. The study tackles the significant problem of fraud by clicking in online advertising, to develop an effective detection technique. It presents an ensemble model for the classification and extraction of features that blend a CNN with (BiLSTM). The ensemble design provides excellent precision (99.89%), recall (98.50%), F1- (99.19%), and specificity (99.89%). This method outperforms existing ensemble and traditional models, which makes it a dependable alternative for protecting pay-per-click advertisements against fraud via clicks and promoting secure online advertising for goods [44].

Subsequent sections are discussed as follows. **Section III** presents an outline of the research process's flow. **Section IV** focuses on analyzing data and visualization charts. **Section V** is completely on Literature Review Synthesis. **Section VI** goes into the metrics used to assess the performance of these models. **Section VII** is devoted to describing the Deep Learning models used in the study. **Section VIII** presents and discusses the model outcomes. Finally, **Section IX** conclusions the article and discussing prospective future research directions.

III. A FLOWCHART OF RESEARCH PROGRESSION

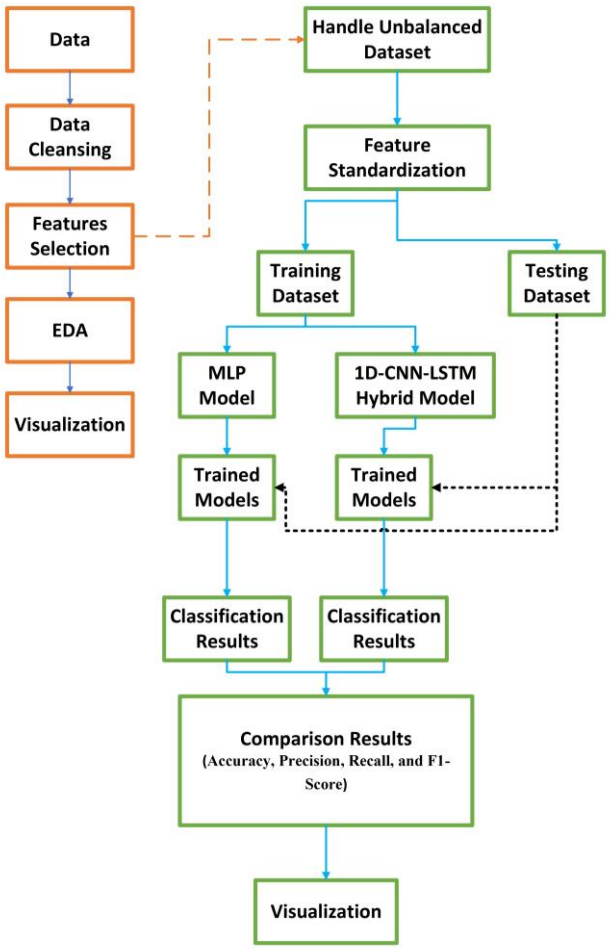

Figure 1: Visualization of a Holistic Workflow Incorporating Data Mining and AI Techniques for Improved Data Science

The investigational methodology of the research project is characterized by a thorough and systematic approach to data science, leveraging a combination of artificial intelligence (AI) and data mining practices. Data purification and selecting features are the initial stages, laying the foundation for transforming data of the highest standard. Through the use of advanced visualization approaches, EDA (exploratory data analysis) can reveal previously undetected trends and provide a better understanding of the data. An essential first step in guaranteeing the model's effectiveness and unbiasedness involves handling unbalanced datasets. The "Testing Dataset" and "Training Dataset" pathways showed a commitment to thorough creation of models and evaluation. The utilization of the "Training Dataset" led to the development of a novel ANN (artificial neural network) and the proposed hybrid deep learning model, reflecting the study's focus on state-of-the-art artificial intelligence (AI) methods. On the other hand, the "Testing Dataset" produced useful "Classification Outcomes" and "Comparative analysis Outcomes", resulting in an accurate

evaluation of the "Trained Models." This thorough and systematic approach combines AI and data mining, producing a framework for study that tackles complicated issues while highlighting the value of simple outcome visualization for increased comprehension and applicability in a variety of environments.

IV. Interpretation of Data and Data Visualization

The data being looked at is a fictitious combination of transactions made with credit cards from January 1, 2019, to December 31, 2020, encompassing both legitimate and fraudulent purchases. It is composed of transactional data from one thousand customers who dealt with eight hundred different retailers. The Sparkov Data Generation tool, developed by Brandon Harris, was utilized to carry out the simulation. The "faker" Python package was utilized in the simulation process to provide a temporary list of preconfigured retailers, customers, and transaction kinds. Distribution of transactions based on many parameters. The resulting dataset offers an accurate representation of simulation transactions by aggregating all individual transactions into one complete set. However, it is critical to notice that the data set is unbalanced, with fewer occurrences of one class in the 'is fraud' targeted variable. This disparity has a likelihood to bias the classification model toward the majority class, affecting its accuracy. The Synthetic Minority Over-Sampling Technique (SMOTE) may be employed to remedy this issue. SMOTE is a well-known strategy for dealing with unbalanced datasets because it creates synthetic samples from the minority class, rebalancing the class distributions and enhancing the accuracy of the model on the minority class [40].
The dataset under examination covers an extensive variety of financial card transaction-related characteristics. The attributes offer an entire overview of each transaction, providing information about the consumer, the item being bought, and the seller involved. Here's an overview of these attributes:

A. *Attributes:*

- trans_date_trans_time: This specifies the date as well as the time of the transaction.
- cc_num: The credit card information that was utilized to complete the transaction.
- merchant: where the transaction occurred is referred to as the merchant.
- category: This defines the sort of money spent (e.g., groceries, shopping).
- amt: The total amount of money exchanged.
- first and last: These are the cardholder's first and last names.
- Gender: This reveals the gender of the cardholder.
- street, city, state, and zip: These provide data about the cardholder's address.
- lat and long: These are the cardholder's location coordinates (latitude and longitude).
- city_pop: the estimated population of the place in which the cardholder lives.
- job: This indicates the cardholder's profession.
- dob: This is the cardholder's date of birth.
- trans_num: This is the transaction's unique identifier.
- unix_time: This is the transaction's time as expressed in Unix format.
- merch_lat and merch_long: These variables are the merchant's location coordinates (latitude and longitude).
- is fraud: This is a feature that contains a binary value that indicates the likelihood that the transaction is non-fraudulent (0) or fraudulent (1). It is used in fraudulent detection algorithms as the target variable.

trans_date_trans_time	zip	lat	long	city_pop	category_code	gender_code	job_code	year_code	month	day	hour	minute	is_fraud
2020-12-31 23:59:07	63453	40.4931	-91.8912	519	0	1	477	1	12	31	23	7	0
2020-12-31 23:59:09	77566	29.0393	-95.4401	28739	5	1	207	1	12	31	23	9	0
2020-12-31 23:59:15	99323	46.1966	-118.9017	3684	5	0	307	1	12	31	23	15	0
2020-12-31 23:59:24	83643	44.6255	-116.4493	129	4	1	63	1	12	31	23	24	0
2020-12-31 23:59:34	73034	35.6665	-97.4798	116001	0	1	289	1	12	31	23	34	0

Figure 2: Top 5 records of the Dataset

The (SMOTE) also known as Synthetic Minority Over-Sampling Method is an essential technique for managing datasets that are imbalanced where the minority class has substantially fewer instances than the majority of the class. SMOTE works by taking samples from the minority class and constructing synthetic examples in the feature environment, thereby oversampling, and rebalancing the dataset [40]. The main steps entail identifying a minority situation and choosing it. The k most nearby neighbors, as well as generating synthetic instances by interpolating between the chosen instance and the closest neighbors [39].

$$\text{Synthetic Instance} = A + \lambda\left(\text{Nearest}_{\text{Neighbor}} - A\right) \quad (1)$$

A representation of the minority instances and ***λ*** is an arbitrary number around 0 and 1 for each of the attributes. This method is especially beneficial in fraud detection historical data, where the rate of fake transactions is typically significantly lower compared to the number of real transactions. SMOTE helps alleviate the class imbalance inherent in such datasets, providing the model with an improved and well-balanced training set. SMOTE adds to the robustness and accuracy of artificial intelligence models in detecting fraudulent activity by retaining the variety of the minority class and providing synthetic examples representative of the actual distribution.

B. *Visualization*

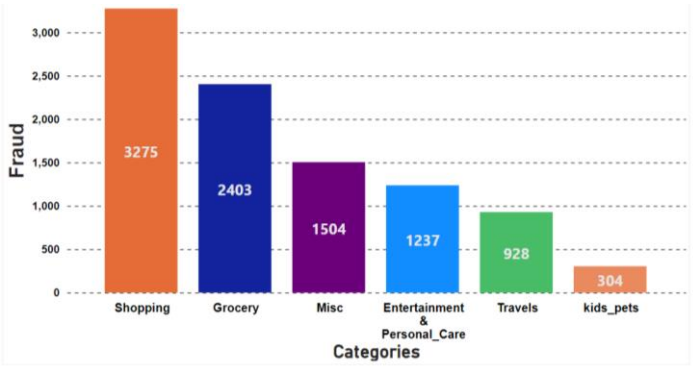

Figure 3: A Detailed Bar Chart Analysis of Fraud Category Distribution

A bar chart depicts a thorough breakdown of fraud types. The chart's X-axis delineates particular fraud types, providing a full

understanding of the types of fraudulent operations. The figure depicts the matching counts of fraud incidents within every category on the Y-axis, demonstrating the prevalence of various forms of fraud. Particularly, the category with the highest incidence is "Shopping," with 3,275 fraud incidents recorded. "Grocery" comes in second with 2,403 instances, highlighting the importance of this area in the larger picture of fraud detection. "Travels" as well as "Kids_Pets" contributed to 928 and 304 cases, respectively, while "Misc" and "Entertainment & Personal_Care" have significant incidences as well, with 1,504 and 1,237 cases, accordingly. This bar chart presents the distribution of fraud across several categories in an eye-catching and informative way, offering those who are interested in recognizing and stopping fraudulent activities an important point of reference.

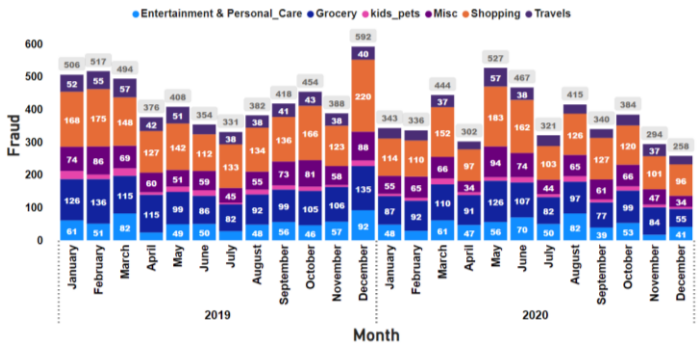

Figure 4: Fraud Detection's Temporal Evolution: A Stacked Column Chart Providing Category-Specific Counts Across Months and Years

The stacked column chart provided a detailed look at the detection of fraud across months and years, thereby providing an improved understanding of the temporal trends and patterns of fraud across various categories. The timeline is represented by the chart's X-axis, with "Months" and "Year" breakdowns giving a simple and thorough reference for monitoring the progress of frauds over time. This temporal divide makes it feasible to spot trends, seasonality, and likely irregularities in the data. The illustration depicts the number of fraud events split across numerous groups on the Y-axis. The categories are shown graphically in different segments, and the height of each segment connects to the number of fraud instances within that group. This stacking of categories within each time frame enables an instantaneous evaluation of the dataset's relative quantity and breakdown of fraud. It highlights what categories have a stronger influence along with how this presence fluctuates as time passes.

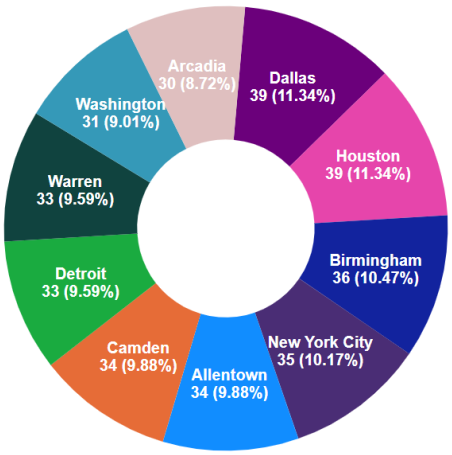

Figure 5: Donut Chart Visualization of the Top 10 Cities with Fraud Incidents

The Donut Chart above presents a concise yet informative picture of the top ten cities in terms of fraud incidents, as measured by the number of "is fraud" instances. Additionally, the data shows that Dallas has the largest number of instances of fraud, with 39 fraud incidents. Houston followed with 39 examples, underscoring its importance in the wider context of identifying fraudulent activity. Birmingham, New York City, and Allentown also have noteworthy occurrences, with 36, 35, and 34 cases, respectively. The size of each part in the donut diagram correlates to the number of fraud incidents thereby showing the corresponding incidence rate of fraud among various cities. The picture gives an in-depth visual tour of the cities with the biggest amounts of fraudulent transactions, which may be particularly beneficial for finding fraudulent activity and risk reduction evaluations in these specific places.

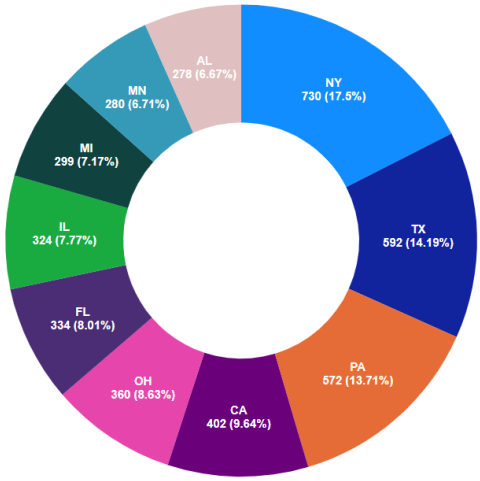

Figure 6: Donut Chart Analysis of the Geographic Breakdown of Fraud incidents in the Top Ten U.S. States

The donut chart illustrates the breakdown of fraudulent activity among the top ten states, providing a short yet complete picture of the fraud environment. Particularly, New York (NY) appears as the state with the greatest level of fraud incidences, with 730 incidents reported. Texas (TX) and Pennsylvania (PA) are close behind, with 592 and 572 instances, respectively. The prominence of these states in the table highlights their vulnerability to fraud. In addition, California (CA) and Ohio (OH) had large fraud numbers of 402 and 360, respectively. The use of state abbreviations simplifies the presentation, allowing for quick recognition and understanding. This visual depiction is a useful tool for researchers and prevention of fraud practitioners, aiding them in the allocation of resources and the development of state-specific methods for more successful prevention and intervention initiatives.

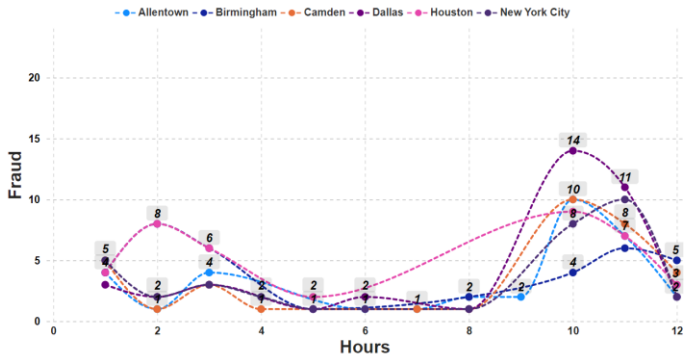

Figure 7: Hourly patterns of identified credit card fraud instances in six major U.S. states on a typical day

The given graph depicts trends of detected fraud cases in 12 hours of a typical day. The observed frauds were noticed in six different states Allentown, Birmingham, Camden, Dallas, Houston, and New York city. Superficially, the highest frauds are detected around 10 AM in the majority of the states except Birmingham. On the other side, all states except Houston have suffered around 2 AM. From 4 AM to 8 AM least fraud cases surfaced in almost all states with minor variations. Particularly, cases started to get observed around 1 AM. During this time frame, almost 3 to 5 cases were accounted for in all states on average. Surprisingly, this phenomenon declined for the next few hours for all states except Houston and Birmingham, which showed a rising trend and reached up to 8 cases each. From 2 AM onward up to 8 AM the rise and fall of detection cases remains almost the same, within the range of 2-4 on average. However, this trend is again varied for the state of Houston as cases suddenly soar upward from 5 AM to 8 AM. Now the most critical time frame is between 8 AM to 12 noon in which cases for all states reached their pinnacle. Especially around 10 AM 14 cases were found in Dallas followed by Camden and Allentown having around 10 fraud cases each. In this regard, Houston and New York city are also not lagging as they have just 1 to 2 cases less during the same period. However, this time zone appears to be safest for the state of Birmingham because merely 4 cases were surfaced and this trend same until the last hour (12noon) for Birmingham state. But, in the remaining states, it was experienced that there is a plunge from 10 AM to 12noon. The
cases for these states ultimately appeared to be in the range of 2-3 cases on average.

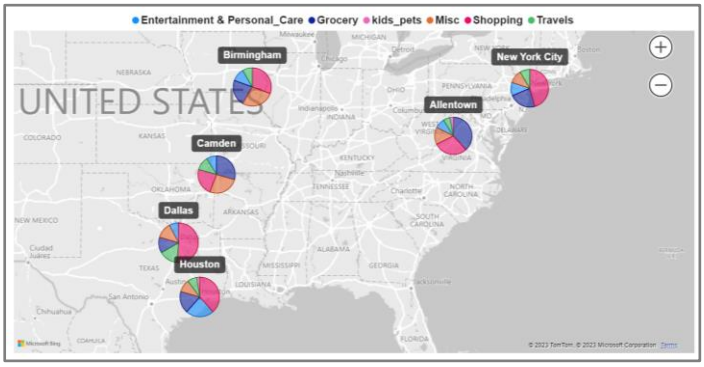

Figure 8: An examination of credit card thefts in six major U.S. states, with an emphasis on the rise in shopping sector fraudulent activities in New York, Dallas, and Houston

The given map illustrates the credit-card fraud detection cases observed in six major states of the USA. The detected frauds are represented through pie-charts in which five different domains were reflected along with some miscellaneous fields. These domains encompass Entertainment & and personal care, grocery, kids-pets, shopping, and travel sectors respectively. Overall, it can easily be visualized that the majority of subject fraud cases were encountered in the shopping sector in the states of New-York city, followed by Dallas and Houston in descending order. Whereas most of the credit card frauds related to grocery were observed in Allentown, Camden, Birmingham, and New York states. The highest number of cases in the Entertainment and personal care sector were observed only in the state of Houston. Considering each state on an individual basis, Houston faced most of the cases in the shopping, Entertainment personal care, and grocery sectors. Whereas the remaining sectors couldn't be regarded as major contributors. As far as the state of Dallas is concerned, more than 60% of issues were caught in only shopping sector. However, all remaining domains have almost shared 10% on average. Camden and Birmingham showed approximately similar pictures regarding the contribution of cases from diverse fields. Namely, Grocery, shopping, and miscellaneous sectors are major victims of credit card fraud in both states and almost 20% of the remaining cases could be accounted for in other sectors. Allentown reflected an alarming picture of cases where the miscellaneous sectors, Grocery, and shopping domains were highly affected by said fraud cases. Eventually, the most densely populated state of New-York had also suffered a lot in terms of fraud in the shopping and Grocery sectors. Conclusively, all states have been victimized in different domains. However, the sectors of Grocery, Entertainment & personal care along shopping were the highest affected among all under the study domains.

V. LITERATURE REVIEW SYNTHESIS

This section tabulates some of the important techniques employed and key findings of other researchers in respective studies, accordingly:

Table 1: Literature Synthesis

Reference	Methodology	Year	Focus	Key Findings
[1]	LightGBM	2023	Wallet-based Transaction Fraud Prevention	Wallet-based transaction fraud prevention through LightGBM with a focus on minimizing false alarms
[6]	Data Mining Techniques	2023	Loan Fraud Detection	Review of loan fraud detection process in the banking sector using data mining techniques
[15]	Explainable Artificial Intelligence	2023	Financial Fraud Detection	Introduced a user-centered explainable AI approach for financial fraud detection
[19]	Role of Artificial Intelligence	2023	Scalable Visual Data Analytics	Investigated the role of artificial intelligence in

				scalable visual data analytics
[21]	Real-time Architecture	2023	Fraud Detection in Online Digital Transactions	Proposed end-to-end real-time architecture for fraud detection in online digital transactions
[29]	Data Analysis and Visualization Techniques	2023	Mobile Money Fraud Detection	Detection of mobile money fraud using data analysis and visualization techniques
[30]	Data Mining Classification	2023	Online Fraud Victim Profile	Compared data mining classification for online fraud victim profiles in Indonesia
[36]	Deep Learning Techniques	2023	Financial Fraud Detection	Used deep learning techniques for financial fraud detection
[39]	Improved Variational Autoencoder GAN	2023	Credit Card Fraud Detection	Improved performance using VAE GAN
[48]	Mixed Method Approach	2023	Big Data Applications in E-commerce	Evaluating large-scale data applications analytically for e-commerce
[57]	Spatial-Temporal-Aware Graph Transformer	2023	Transaction Fraud Detection	Transaction fraud detection via spatial-temporal-aware graph transformer
[67]	Neural Factorization Autoencoder	2023	Online Telephony Fraud Detection	Introduced NFA, a neural factorization autoencoder for online telephony fraud detection
[2]	Machine Learning Algorithms and Quantum Annealing	2022	Online Fraud Detection	Integrated machine learning algorithms with quantum annealing solvers for online fraud detection
[3]	Machine and Deep Learning Algorithms	2022	Credit Card Fraud Detection	Used state-of-the-art machine learning and deep learning algorithms for credit card fraud detection
[7]	Related-Party Transaction Knowledge Graph	2022	Financial Fraud Detection	Used related-party transaction knowledge graph for financial fraud detection
[8]	Community identification algorithms, artificial intelligence models, and control rules	2022	Detecting Fraud Effectively	For efficient identification of fraud, control rules, models based on machine learning, and community detection methods should be combined.
[17]	AI Empowered Big Data Analytics	2022	Industrial Applications	Explored the use of AI-empowered big data analytics for industrial applications
[20]	Big Data Analytics	2022	System for Detecting Credit Card Fraud	discovered fraud with credit cards using big data analytics
[24]	Sequential Pattern Mining	2022	Fraudulent Transaction Detection in Online Banking	Sequential pattern mining approach for personalized fraudulent transaction detection in online banking
[25]	Data Mining	2022	Consumer Fraud in Online Shopping	Data mining was implemented to identify indicators of risk for fraud among consumers in online purchasing.
[28]	Visualization Tool	2022	Financial Fraud Detection	A Visualize Tool for the Identification of Financial Fraud, or ATOVis
[31]	Naïve Bayes Based Classifier	2022	Credit Card Fraud Discovery	Applying a Naïve Bayes classifier to find fraud with credit cards
[32]	Machine Learning and Deep Learning Techniques	2022	Fraud Detection During Financial Transactions	Used machine learning and deep learning techniques for fraud detection during financial transactions
[33]	Machine Learning Algorithms	2022	Fraud Detection in Online Payment Transactions	Fraud detection in online payment transactions using machine learning algorithms

[37]	Combine Deep Learning with Hybrid Optimization	2022	Transactions in Fraud at the Bank	Utilized deep learning and hybrid optimization to identify bogus transactions in the financial institution
[38]	XGBoost Model	2022	Predicting Fraud in Financial Payment Services	Identified fraud associated with payment services using an XGBoost model with improved hyperparameters
[40]	Neural Network Ensemble with Feature Engineering	2022	Credit Card Fraud Detection	Enhanced detection through ensemble and feature engineering
[41]	GAN	2022	Credit Card Fraud Detection	Research on credit card fraud detection based on GAN
[44]	Deep Learning-Based Ensemble Architecture	2022	Click Fraud Detection	Applied ensemble approach for click fraud detection
[45]	Ensemble Model	2022	Credit Fraud Detection	Automatic fraud detection using an ensemble model
[51]	Visualization Techniques	2022	Credit Card Fraud and Money Laundering	Application of visualization methods in the fields of money laundering and fraud with credit cards
[52]	Machine and Deep Learning Techniques	2022	Credit Card Fraud Detection	Credit card fraud detection using machine and deep learning techniques
[53]	Review	2022	Credit Card Fraud Detection Using Machine Learning Techniques	Review on fraud detection in credit card transactions using machine learning techniques
[58]	Literature Review	2022	E-commerce Fraud Detection	Thorough evaluation of the literature on e-commerce preventing and identifying fraud
[63]	Deep Belief Network and Explainable AI	2022	Financial Fraud Detection	DBNex: Deep Belief Network and Explainable AI-based financial fraud detection
[65]	Deep Learning-Based Automated Learning Environment	2022	Corporate Marketing, Business Strategies, Fraud Detection, Financial Time Series Forecasting	Used deep learning for automated learning environments in various business aspects
[66]	Utilization Prediction Technique and Data Mining Architecture	2022	Not Specified	Presented utilization prediction technique and analyzed data mining architecture
[4]	Optimized Convolutional Neural Network	2021	Fake News Detection	Optimized Convolutional Neural Network for fake news detection
[10]	Machine Learning Techniques	2021	Credit Card Fraud Detection System	Novel approach in credit card fraud detection system using machine learning techniques
[16]	ROC Curve-based Thresholding Methods	2021	Online Transactions Fraud Detection	Compared ROC curve-based thresholding methods in online transactions fraud detection systems using deep learning
[26]	Pattern Analysis	2021	Transaction Fraud Detection	Utilized pattern analysis for transaction fraud detection
[62]	Fused Machine Learning	2021	Shill Bidding Fraud Detection	Real-time shill bidding fraud detection empowered with fused machine learning
[14]	Machine Learning Algorithms (QDA, LR, SVM)	2020	Credit Card Fraud Detection	Compared QDA, LR, and SVM for credit card fraud detection
[55]	Graph Neural Network	2020	Fraud Detection via Spatial-temporal Attention	Utilized graph neural network for fraud detection via spatial-temporal attention
[59]	Autoencoder	2020	Credit Card Fraud Detection	Autoencoder-based model for detecting

				fraudulent credit card transactions
[60]	Support Vector Machine	2020	Credit Card Risk Identification	Used selection features and support vector machine for credit card risk identification
[61]	Light Gradient Boosting Machine	2020	Credit Card Fraud Detection	Intelligent approach using an optimized Light Gradient Boosting Machine for credit card fraud detection
[11]	Real-time Credit Card Fraud Detection	2019	Machine Learning	Real-time credit card fraud detection using machine learning
[13]	Big Data Analytics	2019	Banking	Explored the role of big data analytics in banking
[22]	Big Data Analytics	2019	Online E-Commerce Fraud Detection	Scalable approach for fraud detection with big data analytics
[35]	Hybrid Ensemble and Deep Learning	2019	Credit Card Fraud Detection	Champion-challenger analysis using hybrid ensemble and deep learning
[47]	Big Data Analytics in Banking	2019	Role of Big Data Analytics	Explored the role of big data analytics in banking
[54]	Fusion Classifiers	2019	Credit Card Fraud Detection	Analysis of credit card fraud detection using fusion classifiers
[18]	Big Data Analytics Services	2018	Business Intelligence	Explored big data analytics services for enhancing business intelligence
[42]	Generative Adversarial Network	2018	Telecom Fraud Detection	GAN-based approach for telecom fraud detection
[43]	Deep Learning (Auto-Encoder, Restricted Boltzmann Machine)	2018	Credit Card Fraud Detection	Utilized deep learning for fraud detection
[46]	AdaBoost and Majority Voting	2018	Credit Card Fraud Detection	Used AdaBoost and majority voting for detection
[9]	Data Analysis	2017	Fraud Detection	Emphasized fraud detection using data analysis
[12]	Big Data Analytics and Self-Organizing Maps	2017	Online Fraud Detection	User pattern-based online fraud detection and prevention using big data analytics and self-organizing maps
[23]	Data Mining	2017	Credit-Card Fraud Detection in E-tail	Data mining-based system for credit-card fraud detection in e-tail
[27]	Visualization Techniques	2012	Fraud Detection and Behavior Identification	Implemented visualization techniques for fraud detection and behavior identification in online tests
[49]	Overview of Big Data	2011	General	Described big data as the next frontier for innovation, competition, and productivity

VI. CLASSIFICATION MODEL PERFORMANCE METRICS

Confusion Matrix:

A confusion matrix is a table used in classification to evaluate the performance of a model. It compares the predicted classifications to the actual classifications [23].

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

True Positive (TP):
True Positive refers to the number of instances where the model correctly predicted the positive class. In other words, it is the number of observations that are positive and were correctly identified as positive by the model.

False Negative (FN):
False Negative represents the number of instances where the model incorrectly predicted the negative class while the true class is positive. In simpler terms, it is the number of observations that are positive but were mistakenly classified as negative by the model.

False Positive (FP):
False Positive is the number of instances where the model incorrectly predicted the positive class when the true class is negative. It is the number of observations that are negative but were erroneously classified as positive by the model.

True Negative (TN):

The number of times the model correctly forecasts the negative class is known as the True Negative. It is the quantity of negative data that the model properly classified as negative.

Precision:
Precision is the ratio of correctly predicted positive observations to the total predicted positives. It indicates the accuracy of positive predictions.

$$\text{Specificity} = \text{Precision (PR)} = \frac{T_p}{T_p + F_p} \quad (2)$$

Recall:
The ratio of all actual positive results to all accurately anticipated positive observations is known as recall. It measures the ability of the model to capture all the relevant cases.

$$\text{Sensitivity} = \text{Recall (RE)} = \frac{T_p}{T_p + F_n} \quad (3)$$

Accuracy:
Accuracy is the ratio of correctly predicted observations to the total observations. It provides an overall measure of how well the model performs.

$$\text{Accuracy (ACC)} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (4)$$

F1-Score:
F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when there is an imbalance between classes.

$$\text{F1Score} = \frac{2(\text{RE})(\text{PR})}{\text{RE} + \text{PR}} \quad (5)$$

VII. MODEL EXPLANATION

A. *MULTI-LAYER PERCEPTRON:*

Deep learning is built on Artificial Neural Networks (ANNs), which offer an adaptable framework for handling difficult challenges. These models, which include several layers of connected nodes facilitating the flow of data, are modeled after the biological brain. The initial layer, referred to as the input layer, compiles the properties of the data that arrive into a common structure. To indicate the significance of the linkages, weights are applied to node interconnections. Between the layers of input and output, where the data is carefully assessed, ANNs usually consist of hidden layers. To generate non-linearities within these layers—which are essential for identifying complex patterns in the data—a weighted sum of the inputs, biases, and activation functions, such as sigmoid or ReLU, is used. The ability to capture detailed relationships between attributes is enhanced by the use of activation functions. The versatility of the model is increased with functions like sigmoid and ReLU, which enable the network to identify and express complicated patterns. The network's output, which is appropriate for the task at hand—regression or classification—is produced by the last layer, which is specified by an appropriate activation function. Regression tasks could benefit from a linear activation, while many tasks related to classification need activation functions like sigmoid for binary classification or softmax for several classes. ANNs are helpful tools for a variety of machine-learning applications because they work by transmitting data through layers of connected nodes, modifying, and acquiring knowledge from inputs via weighted links, activations, and hidden layers [43].

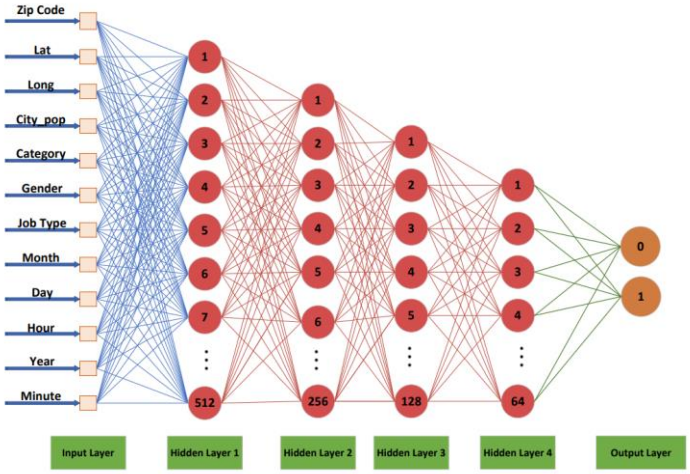

Figure 9: MLP Model Architecture

The weighted sum of a node's inputs in layer i is determined as follows:

$$z_j^{(i)} = \sum_{k-1}^{m^{(i-1)}} W_{jk}^{(i)} a_k^{(i-1)} + b_j^{(i)} \quad (6)$$

where $m^{(i-1)}$ is the number of nodes in layer $i-1, W_{jk}^{(i)}$ is the weight linking node k in layer $i-1$ to node j in layer $i, a_k^{(i-1)}$ is the activation of node k in layer $i-1$, and $b_j^{(i)}$ is the bias of node j in layer i. A node's output after applying the activation function is presented by

$$a_j^{(i)} = \sigma\left(z_j^{(i)}\right) \quad (7)$$

The output of node j after applying the activation function is represented as $a_j^{(i)}$, which is the outcome of using $\sigma\left(z_j^{(i)}\right)$ where $z_j^{(i)}$ is the weighted sum. The feed-forward approach refers to transferring the input throughout the network to reach the ultimate conclusion, and then continuing this procedure for each layer until the desired result is received.

$$a_j^{(i)} = \sigma\left(\sum_{k-1}^{m^{(i-1)}} W_{jk}^{(i)} a_k^{(i-1)} + b_j^{(i)}\right) \quad (8)$$

The function is known as loss which calculates the distance between the model predicted and the real result. The loss function utilized depends on the job at hand (e.g., mse for regression, and however, cross-entropy for classification).

$$L\left(Y, \hat{Y}\right) \quad (9)$$

When establishing a neural network's ultimate conclusion for an input, these mathematical equations are significant. The activation function (σ) enhances the model's degree of non-linear nature and complexity. For example, a sigmoid may compress a result between 0 and 1, reflecting the risk of fraud. The activation function shows how the network behaves to the

various data trends. Accurately capturing and analyzing these patterns is vital for establishing a reliable and efficient model for identifying fraudulent activity. Understanding and selecting the activation function therefore represents an essential part of constructing a neural network for recognizing fraudulent behavior.

B. 1D-CNN-LSTM Proposed Hybrid Model:

In today's sophisticated digital circumstances, powerful artificial intelligence models that are capable of recognizing unseen patterns in sequential data have become crucial for identifying fraud. Linking 1DCNN with the LSTM model is one effective approach to overcome this challenge. This hybrid model, known as a 1D-CNN with LSTM hybrid model, is a good strategy for discovering fraudulent activities, and gathering both local patterns and long-term correlations in the sequential data [43].

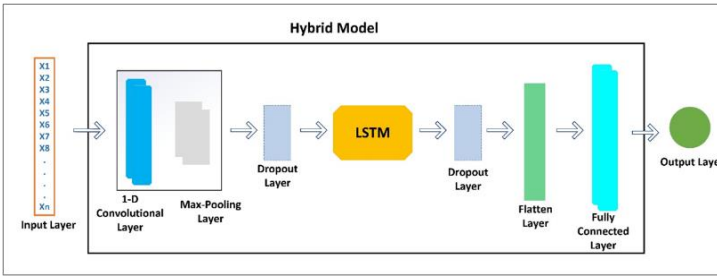

Figure 10: 1D-CNN-LSTM Proposed Model Architecture

The 1D-CNN layer is crucial for finding particular characteristics or patterns in data that are sequential. Local patterns in data collected in time-series format may show separate sequences of transactions, anomalous patterns, or anomalies in fraud detection scenarios. The 1D-CNN's convolutional layers are good at developing and recognizing these localized properties [40].

$$O_i = \sigma\left(\sum_{m-1}^{M} W_m \cdot I_{i+m} + b\right) \quad (10)$$

O_i denotes the output of a neuron at position i in the 1D-CNN layer. $\sum_{m-1}^{M} W_m \cdot I_{i+m}$ captures the convolutional operation. $\boldsymbol{W_m}$ denotes the weights, while $\boldsymbol{I_{i+m}}$ refers to the input data at locations $i + m$ in the input series. The biased factor **b** has been added to the convolutional summation. The bias helps the model adjust for errors or variations in data points that may not be represented by the weights alone. The complete sum is sent via the activation function $\boldsymbol{\sigma}$ are (ReLU) or Sigmoid. The activation function provides a non-linear nature to the model, helping it to learn complicated patterns in the data. Fraudulent activities typically move beyond particular occurrences and can have consequences that last forever. The LSTM section has been developed to capture these prolonged patterns properly. Because LSTMs are great at acquiring and keeping dependencies over extended periods, they are useful for identifying fraudulent activities that involve multiple transactions or happen during a specific temporal sequence. Additionally, by integrating a complicated memory cell architecture, the (LSTM) network, a specific version of a network of recurrent neural networks (RNN), overcomes the issue of replicating sequential input. The LSTM, which is constructed of several connected components, is governed by a collection of equations that enable it to record and preserve long-term correlations in data sequences.

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \quad (11)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (13)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (14)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (16)$$

the forget gate ($\boldsymbol{f_t}$) is essential for deciding which data from a previous point in time should be retained and which should be discarded. A sigmoid that reduces data from zero to one is used to achieve this. When the value is close to 0, it means that the related data is removed, and when it is close to 1, it means that the data is retained. The flow of fresh points of information into the cell state ($\boldsymbol{C_t}$) is then controlled by the input gate ($\boldsymbol{i_t}$). The candidate cell state ($\boldsymbol{\tilde{C}_t}$) can be refreshed in portions that can be identified with the use of the sigmoid activation function. The tanh activation function is used to establish the non-linearity and make sure the values range from -1 to 1. The candidate cell state passes through this function. The model can decide whether to add or remove data from the cell state during this phase. These elements work together to produce the updated cell state ($\boldsymbol{C_t}$), where memory recall as well as retention are simultaneously impacted by the forget gate and input gate. The amount of cell state that is exposed to the next layer is determined by the output gate ($\boldsymbol{o_t}$). The final output of the LSTM cell is a state that is hidden ($\boldsymbol{h_t}$), which is produced by sending this output via the tanh. In the context of all of these equations, the terms "bias" and "weight" have relevance. The model gains an additional degree of freedom from the bias elements ($\boldsymbol{b_f}$, $\boldsymbol{b_i}$, $\boldsymbol{b_C}$, $\boldsymbol{b_o}$), which enables it to correct for data shifts or offsets that the weights alone could ignore. The weights ($\boldsymbol{W_f}$,$\boldsymbol{W_i}$,$\boldsymbol{W_C}$,$\boldsymbol{W_o}$) determine the relative influence of each input on the corresponding gate or cell state, hence affecting the data flow as well as processing throughout the LSTM. Fundamentally, the forgetting gates, input gates, output gates, biases, and weights included in the LSTM's design enable it to retain and refresh data intelligently, hence enabling the capturing of temporal correlations. This ability is particularly helpful in scenarios like detecting fraudulent activity when it's vital to identify both long-term trends and short-term anomalies.

VIII. RESULT AND DISCUSSION

In this portion of the paper, we examine the outcomes of two distinct models designed for the essential job of fraud detection. The occurrence of fraudulent transactions. The metrics have been selected to offer a multidimensional perspective of the algorithms' ability to identify fraud. The (MLP) and an innovative 1D-CNN_LSTM Hybrid model, both of which are rigorously trained throughout 100 epochs with a batch size of 512, are models under evaluation in Binary Accuracy. During the process of the optimization stage, the models are evaluated using a binary-cross-entropy loss function. Model parameters in both the MLP and 1D-CNN_LSTM Hybrid models are maintained identically to ensure a fair and comparative evaluation. The

selection of assessment measures, particularly Binary Accuracy, Precision, F1-score, and Recall, is extremely important in the context of fraud classification. Furthermore, the decision-making threshold for accuracy, precision, and recall metrics has been set at 0.5, reflecting the binary classification aspect of identifying fraudulent transactions. The metrics are chosen to offer a multidimensional perspective of the algorithms' ability to identify fraud. Precision evaluates the models' ability to minimize false positives, which can be vital in assessing the context of fraudulent conduct. Recall points out the models' capacity to recognize true events of fraudulent activity, and the F1-score maintains the right balance between Precision and Recall.

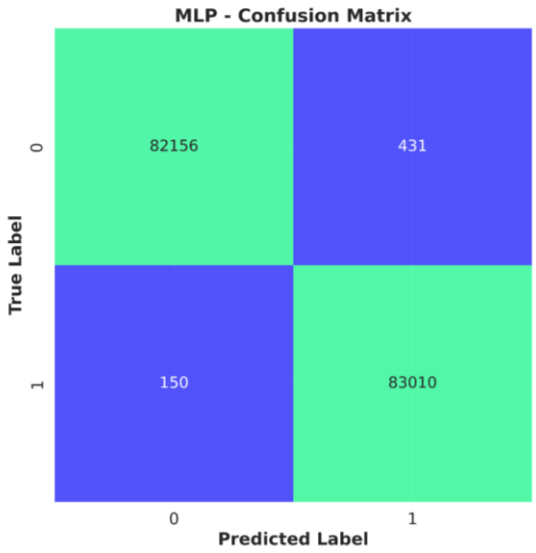

Figure 11: Confusion Matrix for MLP Model in Fraud Detection

The MLP (Multilayer Perceptron) model has been trained, and the performance metrics provide reliable fraud detection results. The training set of metric accuracy of 99.44% reveals the model's ability to provide appropriate classifications. The model's ability to identify genuine positives while reducing false positives is demonstrated by its precision of 99.25%. A fair trade-off between recall and precision is represented by the F1 score of 99.44%, which illustrates that the model can catch both false positives and false negatives. Moreover, the 99.63% recall score demonstrates the model's capacity to identify instances of genuine fraud. With an accuracy of 99.64%, the MLP model still performs brilliantly after validation. Maintaining 99.47%, the model's accuracy on the validation set of results validates its capacity to lower false positives. The model's balanced performance is further shown by its F1-score of 99.64%, which is a harmonic mix of precision and recall. With a recall score of 99.82% on the validation set, the model's ability to effectively identify real-world fraud situations is confirmed. On the other hand, diagonal dominance, which indicates the model's capacity for accurate classification, is expected to be reflected in the confusion matrix chart that is produced from these observations.

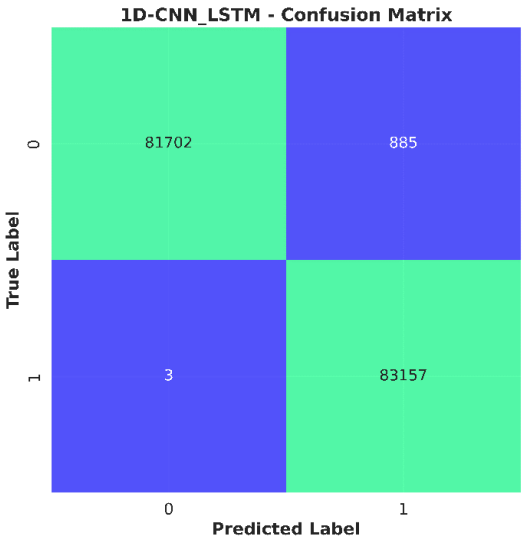

Figure 12: Confusion Matrix for 1D-CNN_LSTM Proposed Model in Fraud Detection

The excellent performance results of the 1D-CNN_LSTM Hybrid model demonstrate its success in identifying fraud. During the training phase, the model reaches an accuracy of 99.56%, indicating its ability to classify data accurately. With a precision of 99.15%, the model shows that it is capable of recognizing true positives while lowering false positives. With accuracy and recall harmonic combined, the model's F1-score of 99.36% shows that it can capture both false positives and false negatives. The recall score of 99.57% reveals how well the approach can recognize instances of real fraud. Following validation, the 1D hybrid model maintains its high level of performance, achieving an accuracy of 99.76%. On the validation set, the accuracy of the model is 98.94%, g of true fraud. The confusion matrix graphic that was produced as a consequence of using CNN_LSTM confirms its capacity to lower false positives. The model's balanced performance is demonstrated by its outstanding 99.46% F1-score, a comprehensive measure that combines accuracy and recall. It is anticipated that diagonal dominance would display the model's accuracy and incorrect classifications when the validation sample recall score reaches an astounding 99.99%, highlighting the model's robustness in identifying instances using these metrics. The 1D-CNN_LSTM superior accuracy and recall scores demonstrate its ability to effectively weed out false positives and false negatives, which translates to its competence in the challenging field of fraud detection.

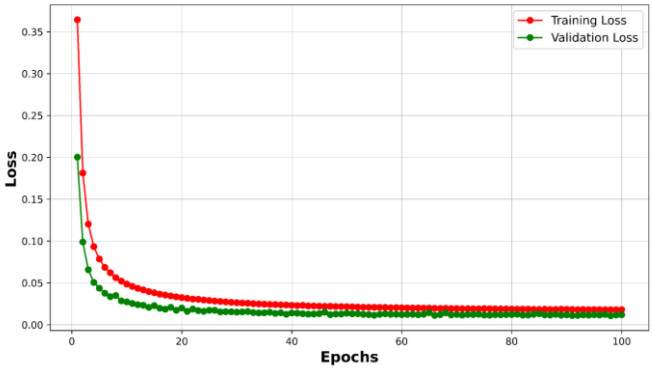

Figure 13: Training and Validation Loss Performance of MLP Model

The training and validation loss graphs of the (MLP) model offer fascinating insights into how it learns. The training and validation losses show a declining trend, indicating that the model is exhibiting excellent convergence. The average error between the actual and expected values on the training set is represented by a loss of training of 0.0181. Concurrently, a second validation dataset repeats this pattern with a validation loss of 0.0120, suggesting that the model generalizes well to data that is not known. The accuracy of the model in identifying authentic transactions from fraudulent ones is 99.44% during training and 99.64% during validation. These results demonstrate how, despite overfitting the training set, the model is robust in detecting the underlying fraud behaviors.

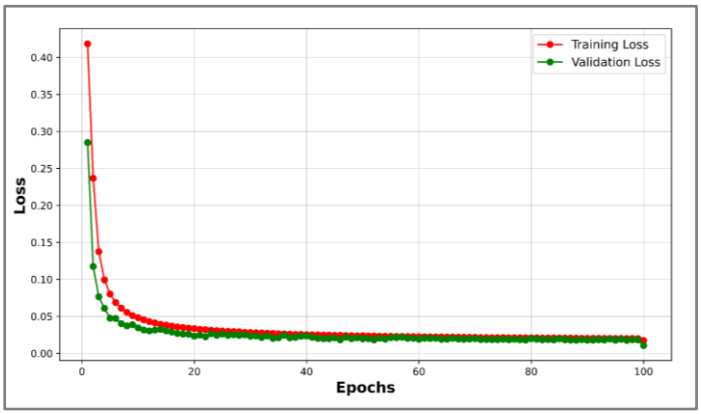

Figure 14: Training and Validation Loss Plot Study for 1D-CNN_LSTM Hybrid Model

The training and validation loss curve for the 1D-CNN_LSTM gives significant insight into how it learns behavior. The training loss of 0.0174 illustrates that the model has the potential to effectively eliminate errors throughout the training period. At the same period the validation loss, which attained 0.0106, illustrates the model's competence in generalizing on not known data, suggesting its capacity to grasp underlying patterns. The 1D-CNN_LSTM distinguishes between real and fake transactions with a training accuracy of 99.56% and a validating accuracy of 99.76%. Furthermore, the training and validation losses, combined with continually outstanding accuracy values, illustrate the 1D CNN_LSTM ability in fraud identification without the overfitting training datasets.

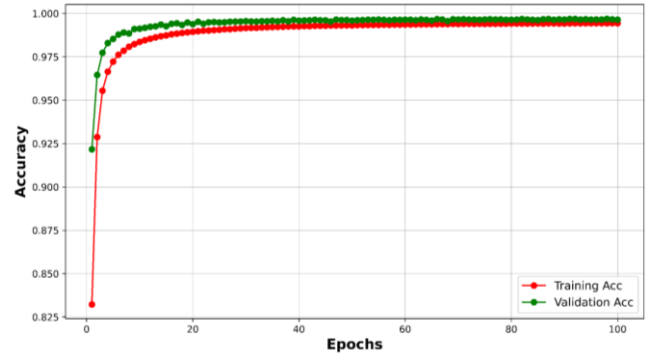

Figure 15: Training and Validation Accuracy Plot Analysis for MLP Model (100 Epochs)

The training and validation accuracy curve of the MLP model showcases a compelling narrative of its learning progression over 100 epochs. Upon completion of this training expedition, the model exhibits a training accuracy of 99.44%, demonstrating its proficiency in precisely classifying occurrences within the training dataset. The increasing accuracy trend demonstrates the model's continuous improvement in identifying the intricate patterns associated with fraudulent and legitimate transactions. As the model gets used to the unfamiliar data that makes up the set of validations, its ability to generalize becomes evident. The model's validation accuracy of 99.64% showcases its ability to effectively utilize previously acquired knowledge on new instances, thereby emphasizing its robustness beyond the training data. The close correspondence between the training and validation accuracies demonstrates that the model not only retains the complicated details of the training set but also effectively applies its knowledge to unfamiliar data. The consistent and elevated level of precision throughout the 100 epochs illustrates the model's reliability in identifying fraudulent activities. The model's ability to effectively learn and adjust to the challenges involved in identifying fraudulent transactions is demonstrated by the consistent increase and alignment of both training and validation accuracy across the course of the epochs.

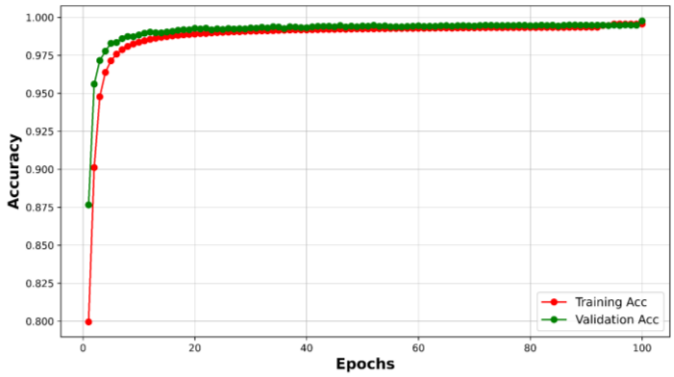

Figure 16: Training and Validation Accuracy Plot Analysis for 1D-CNN_LSTM Hybrid Model (100 Epochs)

The accuracy map of the 1D-CNN_LSTM hybrid model used during training and validation showcases a remarkable illustration of its learning patterns across 100 epochs. After this extended training period, the model obtained an amazing training accuracy of 99.56%, indicating its ability to accurately classify instances within the training dataset. The enhanced accuracy demonstrates the continuous advancement of the model as it acquires complex patterns that are crucial for effectively identifying fraudulent transactions. The validation accuracy of 99.76% reveals the model's potential to transfer its learned information to fresh instances, pointing to its ability to adapt beyond the training data. The close alignment of training and validation accuracy levels implies not just an understanding of the training set difficulties, but also effective generalization to novel, previously observed data.

Table 2: Model Performance Metrics in Fraud Detection Research in Different Areas

Study	Area	Model	Acc	Precision	Recall	F1-Score
[2]	Credit Card Fraud Detection	SVM	-	0.8100	0.6310	-
		MLP	-	0.8410	0.6770	-

		Decision Tree	-	0.7900	0.6250	-
		AdaBoost	-	0.8800	0.7080	-
		LSTM	-	0.8490	0.6740	-
[53]	Automated Insurance Systems: Fraud Detection and Risk Measurement	Decision Tree	0.7444	0.6473	0.5953	0.6005
		SVM	0.7321	0.6696	0.5652	0.4841
		Nearest Neighbor	0.7380	0.6696	0.5256	0.4841
		XGBoost	0.7681	0.6828	0.6295	0.6392
[4]	Click Fraud Detection in Pay-Per-Click Advertisement Campaign	BiLSTM	0.9627	0.9650	0.9630	0.9630
		CNN	0.9613	0.9630	0.9610	0.9610
		CNN-BiLSTM	0.9941	0.9940	0.9940	0.9940
		CNN-BiLSTM-RF	0.9958	0.9960	0.9960	0.9942
OUR	**Fraudulent Transaction Detection**	**MLP**	**0.9964**	**0.9947**	**0.9963**	**0.9964**
		1D-CNN-LSTM	**0.9976**	**0.9894**	**0.9957**	**0.9946**

Valuable insights into the performance of each model in automated insurance platforms and the identification of fraud may be gained from comparing them across research studies. In studies on credit card fraud detection, several models—including SVM, MLP, Decision Tree, AdaBoost, and LSTM—display different capacities [2]. While each model succeeds in different parameters such as precision, recall, and accuracy, it is clear that the LSTM model shines out with a precision of 0.8490 and a recall of 0. 6740. Furthermore, the decision tree, SVM, nearest neighbor, and XGBoost models are examined in the Automated Insurance Systems research [53]. In this case, XGBoost stands out as a strong performer, with the greatest accuracy of 0.7681 and remarkable precision and F1-score values. However, in the Click Fraud Detection research [4,] models such as BiLSTM, CNN, CNN-BiLSTM, and CNN-BiLSTM-RF exhibit exceptional accuracy, with CNN-BiLSTM-RF winning with 0.9958. Surprisingly, in our study concentrating on Fraudulent Transaction Detection, both MLP and 1D-CNN_LSTM hybrid models perform exceptionally well. The 1D-CNN_LSTM model outperforms the competition with the greatest accuracy of 0.9976, precision of 0.9894, recall of 0.9957, and F1-score of 0.9946, which makes it the best model across several crucial criteria. As a result, when it comes to detecting fraudulent transactions, the 1D-CNN_LSTM model emerges as a highly promising and effective option, providing a complete and balanced performance across precision, recall, and accuracy metrics.

IX. CONCLUSION

In the final analysis, this research looks into the vital arena of fraudulent transaction classification, combining modern data mining and visualization strategies and efficient deep learning models to address the growing danger of financial fraud. The usage of the Multilayer Perceptron (MLP) with the unique 1D Convolutional Neural Network and Long Short-Term Memory (1D-CNN-LSTM) hybrid model displays promising outcomes in accurately recognizing fraudulent activity. This research understands the fundamental issue of unbalanced datasets and successfully minimizes bias with the Synthetic Minority Over-Sampling Technique (SMOTE), delivering a more robust and fair training and assessment procedure. The comparison research demonstrates that even though the MLP model improves in precision and recall, the 1D-CNN-LSTM exceeds in accuracy and recall, providing practitioners with significant insights that correlate model selection with specific application needs. Moreover, the use of data mining for visualization provides a distinct layer to the research, enabling a greater knowledge of the subtle patterns inside fraudulent transactions. The findings add to the ongoing attempts to strengthen fraud detection approaches, helping practitioners to make educated decisions customized to their aims. Additionally, the current study not only highlights the usefulness of the MLP and 1D-CNN-LSTM hybrid models but also underlines the significance of a nuanced approach in picking the most suited model depending on the objectives of a specific application. By shining light on the strengths and limits of these models, the study opens the path for breakthroughs in fraud detection tactics and supports the continuous development of cutting-edge technology in the ongoing war against financial misconduct.

REFERENCES

[1] C. Iscan, O. Kumas, F. P. Akbulut, and A. Akbulut, 'Wallet-based transaction fraud prevention through LightGBM with the focus on minimizing false alarms', *IEEE Access*, pp. 1–1, 2023.

[2] W. Wang, Y. Liu, and B. Alidaee, 'Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection', IEEE Access, vol. 10, pp. 75908–75917, 2022.

[3] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, 'Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms', IEEE Access, vol. 10, pp. 39700–39715, 2022.

[4] H. Saleh, A. Alharbi, and S. H. Alsamhi, 'OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection', IEEE Access, vol. 9, pp. 129471–129489, 2021.

[5] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, 'A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement', IEEE Access, vol. 8, pp. 58546–58558, 2020.

[6] F. S. Esmail, F. K. Alsheref, and A. E. Aboutabl, 'Review of Loan Fraud Detection Process in the Banking Sector Using Data Mining Techniques', International journal of electrical and computer engineering systems, vol. 14, no. 2, pp. 229–239, 2023.

[7] X. Mao, H. Sun, X. Zhu, and J. Li, 'Financial fraud detection using the related-party transaction knowledge graph', Procedia Computer Science, vol. 199, pp. 733–740, 2022.

[8] T. Tadesse, 'Combining Control Rules, Machine Learning Models, and Community Detection Algorithms for Effective Fraud Detection', in 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 2022, pp. 42–46.

[9] R. U and B. S. Babu, 'Real-time credit card fraud detection using Streaming Analytics', in 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 439–444.

[10] G. M. Suhas Jain, N. Rakesh, K. Pranavi, and L. Bale, 'A Novel Approach in Credit Card Fraud Detection System Using Machine Learning Techniques', in 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), 2021, vol. 1, pp. 1–5.

[11] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, 'Real-time Credit Card Fraud Detection Using Machine Learning', in 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019, pp. 488–493.

[12] N. Balasupramanian, B. G. Ephrem, and I. S. Al-Barwani, 'User pattern based online fraud detection and prevention using big data analytics and self organizing maps', 2017 International Conference on Intelligent

Computing, Instrumentation and Control Technologies (ICICICT), 2017.
[13] T. Gupta, N. Gupta, A. Agrawal, A. Agrawal, A. Agrawal, and K. Kansal, 'Role of Big Data Analytics In Banking', International Conferences on Contemporary Computing and Informatics, 2019.
[14] P. Naveen and B. Diwan, 'Relative Analysis of ML Algorithm QDA, LR and SVM for Credit Card Fraud Detection Dataset', in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 976–981.
[15] Y. Zhou, H. Li, Z. Xiao, and J. Qiu, 'A user-centered explainable artificial intelligence approach for financial fraud detection', Finance Research Letters, vol. 58, p. 104309, 2023.
[16] Kanika, J. Singla, and Nikita, 'Comparing ROC Curve based Thresholding Methods in Online Transactions Fraud Detection System using Deep Learning', in 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 9–12.
[17] V. D. A. Kumar, V. Vijayakumar, M. K. Gupta, J. J. P. C. Rodrigues, and N. Janu, 'AI Empowered Big Data Analytics for Industrial Applications', Journal of Universal Computer Science, 2022.
[18] Z. Sun, L. Sun, and K. D. Strang, 'Big Data Analytics Services for Enhancing Business Intelligence', null, 2018.
[19] R. Raman, D. Buddhi, G. Lakhera, Z. Gupta, A. Joshi, and D. Saini, 'An investigation on the role of artificial intelligence in scalable visual data analytics', Artificial Intelligence and Symbolic Computation, 2023.
[20] A. Alshammari, R. Alshammari, M. Altalak, K. Alshammari, and A. Alhakamy, 'Credit-card Fraud Detection System using Big Data Analytics', 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022.
[21] A. Hanae, B. Abdellah, E. Saida, and G. Youssef, 'End-to-End Real-time Architecture for Fraud Detection in Online Digital Transactions', International Journal of Advanced Computer Science and Applications, 2023.
[22] Z. Hangjun et al., 'A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics', Cmc-computers Materials & Continua, 2019.
[23] N. S. Carneiro, G. Figueira, and M. Costa, 'A data mining based system for credit-card fraud detection in e-tail', null, 2017.
[24] J. Kim, H. Jung, and W. Kim, 'Sequential Pattern Mining Approach for Personalized Fraudulent Transaction Detection in Online Banking', Sustainability, 2022.
[25] T. Knuth and D. C. Ahrholdt, 'Consumer Fraud in Online Shopping: Detecting Risk Indicators through Data Mining', International Journal of Electronic Commerce, 2022.
[26] N. Cochrane et al., 'Pattern Analysis for Transaction Fraud Detection', null, 2021.
[27] H. Kalidasu, B. Prasannakumar, and Haripriya, 'A Fraud Detection based Online Test and Behavior Identification Implementing Visualization Techniques', null, 2012.
[28] 'ATOVis—a Visualization Tool for the Detection of Financial Fraud', null, 2022.
[29] R. Al-Sayyed, E. Alhenawi, H. Alazzam, A. Wrikat, and D. Suleiman, 'Mobile money fraud detection using data analysis and visualization techniques', Multimedia tools and applications, 2023.
[30] S. Sunardi, A. Fadlil, and N. M. P. Kusuma, 'Comparing Data Mining Classification for Online Fraud Victim Profile in Indonesia', INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 2023.
[31] R. O. Ogundokun, S. Misra, S. Misra, O. J. Fatigun, and J. K. Adeniyi, 'Naïve Bayes Based Classifier for Credit Card Fraud Discovery', Lecture notes in business information processing, 2022.
[32] M. Y. Turaba, M. Hasan, N. I. Khan, and H. A. Rahman, 'Fraud Detection During Financial Transactions Using Machine Learning and Deep Learning Techniques', International Conference on Communications, Computing, Cybersecurity, and Informatics, 2022.
[33] D. Aladakatti, G. P, A. Kodipalli, and S. Kamal, 'Fraud detection in Online Payment Transaction using Machine Learning Algorithms', 2022 International Conference on Smart and Sustainable Technologies in Energy and Power Sectors (SSTEPS), 2022.
[34] J. Li, 'E-Commerce Fraud Detection Model by Computer Artificial Intelligence Data Mining', Computational Intelligence and Neuroscience.
[35] E. Kim et al., 'Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning', Expert Systems With Applications, 2019.
[36] A. Dileep, A. Karthik, G. S. Krishna, D. Ganesh, and S. Hariharan, 'Financial Fraud Detection Using Deep Learning Techniques', 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023.
[37] C. S. Kolli and U. D. T., 'Hybrid Optimization and Deep Learning for Detecting Fraud Transactions in the Bank', International Journal of Information Security and Privacy, 2022.
[38] S. Dalal, B. Seth, M. Radulescu, C. Secară, and C. Tolea, 'Predicting Fraud in Financial Payment Services through Optimized Hyper-Parameter-Tuned XGBoost Model', Mathematics, 2022.
[39] Y. Ding, W. Kang, J. Feng, B. Peng, and A. Yang, 'Credit Card Fraud Detection Based on Improved Variational Autoencoder Generative Adversarial Network', IEEE Access, 2023.
[40] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, 'A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection', IEEE Access, 2022.
[41] J. Ge, X. Liao, and Y. Fang, 'Research on Credit Card Fraud Detection Based on GAN', International Conference on Robotics, Intelligent Control and Artificial Intelligence, 2022.
[42] Y.-J. Zheng et al., 'Generative adversarial network based telecom fraud detection at the receiving bank', Neural Networks, 2018.
[43] 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, 2018.
[44] A. Batool and Y. Byun, 'An Ensemble Architecture Based on Deep Learning Model for Click Fraud Detection in Pay-Per-Click Advertisement Campaign', IEEE Access, 2022.
[45] S. Srinidhi, K. Sowmya, and S. Karthika, 'Automatic Credit Fraud Detection Using Ensemble Model', *null*, 2022.
[46] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, 'Credit Card Fraud Detection Using AdaBoost and Majority Voting', *IEEE Access*, 2018.
[47] R. E. Cascarino, 'Fraud Detection Using Data Analysis', *null*, 2017.
[48] A. Mohammadi, 'Analytical evaluation of big data applications in E-commerce: A mixed method approach', *Decision Science Letters*, 2023.
[49] J. Manyika, 'Big data: The next frontier for innovation, competition, and productivity', *null*, 2011.
[50] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, 'BankSealer: A decision support system for online banking fraud analysis and investigation', *Computers & Security*, 2015.
[51] M. E. Lokanan, 'Financial fraud detection: the use of visualization techniques in credit card fraud and money laundering domains', *Journal of Money Laundering Control*, 2022.
[52] S. Sharma, A. Kataria, J. K. Sandhu, and K. R. Ramkumar, 'Credit Card Fraud Detection using Machine and Deep Learning Techniques', *2022 3rd International Conference for Emerging Technology (INCET)*, 2022.
[53] T.C. Rosshan, 'REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES', *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 2022.
[54] P. Kumari and S. P. Mishra, 'Analysis of Credit Card Fraud Detection Using Fusion Classifiers', *null*, 2019.
[55] D. Cheng, X. Wang, Y. Zhang, L. Zhang, and L. Zhang, 'Graph Neural Network for Fraud Detection via Spatial-temporal Attention', *IEEE Transactions on Knowledge and Data Engineering*, 2020.
[56] C. S. Kolli, and U. D. Tatavarthi, 'Fraud detection in bank transaction with wrapper model and Harris water optimization-based deep recurrent neural network', *Kybernetes*, 2020.
[57] Y. Tian and G. Liu, 'Transaction Fraud Detection via Spatial-Temporal-Aware Graph Transformer', *arXiv. org*, 2023.
[58] V. F. Rodrigues *et al.*, 'Fraud detection and prevention in e-commerce: A systematic literature review', *Electronic Commerce Research and Applications*, 2022.
[59] S. Misra, S. Thakur, M. Ghosh, S. K. Saha, and S. K. Saha, 'An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction', *Procedia Computer Science*, 2020.
[60] N. Rtayli, N. Enneya, and N. Enneya, 'Selection Features and Support Vector Machine for Credit Card Risk Identification', *Procedia Manufacturing*, 2020.
[61] A. A. Taha and S. J. Malebary, 'An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine', *IEEE Access*, 2020.
[62] W. U. H. Abidi *et al.*, 'Real-Time Shill Bidding Fraud Detection Empowered With Fussed Machine Learning', *IEEE Access*, vol. 9, pp. 113612–113621, 2021.
[63] A. Bhowmik, M. Sannigrahi, D. Chowdhury, A. D. Dwivedi, and R. Rao Mukkamala, 'DBNex: Deep Belief Network and Explainable AI based

Financial Fraud Detection', in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 3033–3042.

[64] T. Gupta, N. Gupta, A. Agrawal, A. Agrawal, and K. Kansal, 'Role of Big Data Analytics In Banking', in *2019 International Conference on contemporary Computing and Informatics (IC3I)*, 2019, pp. 222–227.

[65] Z. Bouzidi, M. Amad, and A. Boudries, 'Deep Learning-Based Automated Learning Environment Using Smart Data to Improve Corporate Marketing, Business Strategies, Fraud Detection in Financial Services, and Financial Time Series Forecasting', in *International Conference on Managing Business Through Web Analytics*, 2022, pp. 353–377.

[66] R. Manocha, E. P. Kaur, and E. N. Dhariwal, 'Utilization Prediction Technique and Analyze Data Mining Architecture', in *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 2022, pp. 110–113.

[67] A. Wahid, M. Msahli, A. Bifet, and G. Memmi, 'NFA: A neural factorization autoencoder based online telephony fraud detection', *Digital Communications and Networks*, 2023.