IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

A Review of Emerging Deep Learning Approaches and Technologies for Fraud Detection in Financial Systems

Mr. Kapil Ahir
Assistant Professor, Department of Computer Sciences and Applications
Mandsaur University, Mandsaur
kapil.ahir@meu.edu.in

Abstract: *Digital transactions, internet banking, and advanced financial networks have all contributed to the rise of financial fraud as a pressing issue in today's financial system. This is a review paper on financial fraud detection and involves details of fraud, data characteristics, challenges of detection, and advanced forms of examination. Critical points such as the imbalance of classes, concept drift, a requirement to find in real-time and privacy constraints are considered. The paper analyzes the data mining methods, including classification, prediction, and regression, and the new methods of deep learning, including Convolutional Neural Networks, Recurrent Neural Networks, Graph Neural Networks, autoencoders, and ensemble models. Furthermore, the popular credit card fraud, financial statement fraud, and synthetic fraud simulation benchmark datasets are addressed in a bid to highlight how they can be applied in the model assessment. The findings highlight the conclusions that deep learning-based and hybrid designs are superior to more conventional measures in order to capture complex trends and relationships, which are typical of fraudulent transactions. The study would be helpful in informing the arguments and practitioners that seek to devise effective, scalable and intelligent financial fraud detection models.*

Keywords: Fraud Detection, Deep Learning, Financial Systems Security, Machine Learning in Finance, Credit Card Fraud Detection, Risk Management Systems

I. INTRODUCTION

The financial sector plays one of the most important roles in attaining economic sustainability [1]. The effective operation of the financial system is very important in economic sustainability by efficient mobilization of funds, risk and information management, promotion of innovation, development of technology, and increased productivity. It may be considered that the process of prediction of the price movement of the trends within the financial markets worldwide, corporate profits, and financial products such as stocks is an incredibly challenging one because it depends on a wide range of complicated factors [2]. In addition to external economic factors, like political events and exchange rates, take into account economic variables like GDP and interest rates, basic and technical indicators [3].

One of the most pressing issues of current business environments is financial fraud, the impact of which may be catastrophic in both individual businesses and the global economy in general [4]. The recent thorough surveys have indicated that 56% of companies that are situated all over the globe have reported fraud of some sort; financial fraud is one of the most common and most economically devastating types of this fraud. Fraud events are complex and large in scale and continue to evolve and complicate the use of the old-fashioned ways of detection, making them say the slightest use to be ineffective in dealing with the current threat.

Digital payment systems are also increasingly used, which has revolutionized the financial landscape [5], as they have become convenient and efficient to conduct transactions. This revolution has seen credit card usage become the order

DOI: 10.48175/IJARSCT-32137

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

of the contemporary business era as one finds it easy to purchase goods and services using the card. Credit card fraud is an ongoing and expensive issue in financial institutions, and it has been escalating with the advent of digital financial solutions [6]. In addition to leading to large-scale losses of money to financial institutions as well as to the issuing authorities of the card, fraudulent transactions result in weakening the integrity of the digital payments, which is a significant challenge to the growth and prosperity of the digital economy.

The decent systems of fraud detection are applicable to increase the security and reliability of payment to the consumers and the financial institutions. The systems assist in curbing illegal business, reducing risks and not betraying financial transactions. This minimizes possible loss of money, increases confidence of customers and is part of the greater objective, which is to have a safe digital financial ecosystem. AI [7] has become a disruptive entity that can revolutionize the process of detecting fraud. Leveraging ML [8] and DL [9], AI-based systems are progressively used to uncover anomalous tendencies in huge data, determine fraudulent activity in real-time, and minimize financial losses. The usefulness of such systems in all industries, such as banking, insurance, and healthcare, can now be discussed and researched extensively. Real-world applications of DL [10] algorithms include financial institutions, insurance firms, and computer networks; they are also employed in intrusion detection systems for mobile cellular networks and healthcare facility monitoring services for medical fraud. They are applied in detection, automation of homes, detection of malware in Android, video surveillance, tracking of locations, medical diagnosis and predicting heart disease amongst others.

A. Structure of the Paper

The paper is organized as follows: Section II covers fraud detection in the financial system, Section III gives a brief discussion on the new types of deep learning, Section IV discusses benchmarking datasets, Section V covers the literature review, and Section VI summarizes the paper and gives it its future research directions.

II. FRAUD DETECTION IN THE FINANCIAL SYSTEM

Financial theft is the act of getting money by sneaking around the law and lying. The financial fraud may be performed in other spheres, including insurance, banking, taxation, and corporate. Financial transaction fraud, money laundering, and other forms of financial fraud have lately turned out to be a growing challenge among businesses and industries. Every day, huge sums of money are wasted due to fraud, and this persists despite several attempts to limit it. The economy and society suffer as a result.

A. Types of Financial Frauds

Financial fraud has received significantly increased attention over the past decade owing to the potential repercussions of undetected irregularities on the industry and daily life. The shape of such crimes may be different, yet they all can disrupt the economies, spike prices, and rattle the consumer trust in shopping. Many forms of deceit fall under that umbrella; one example is bank fraud, which encompasses a variety of scams including mortgage, credit card, and money laundering. Fraudulent claims involving insurance policies for crops, healthcare, vehicles, and other assets are another prevalent kind of financial fraud. Additional forms of financial fraud include insider trading, fraud involving commodities and securities, and others. Figure 1 illustrates the many forms of financial fraud:

Fig. 1. Types of Financial Frauds

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

• Credit Card Frauds: A "credit card" is a type of payment card that a customer (the user) can use to buy things up to their credit limit or pay for them in cash ahead of time [11]. Credit cards give users the benefit of the doubt by pushing back the payment due date to the next billing cycle. This gives them more time to pay off their balances within the required repayment term. Criminals who use credit cards are easy to catch. Silently withdrawing large sums of money without the owner's knowledge is both secure and fast. Detecting fraud is a daunting and perilous endeavor since con artists are persistent in their efforts to pass off fraudulent transactions as legitimate.

• Health Insurance Frauds: Health insurance fraud detection uses past claims data to spot out-of-the-ordinary occurrences and assign relative probabilities of fraud based on those findings [12]. The major goal of fraud detection is to focus investigators' attention on the most likely cases of fraud, as domain specialists need to carry out additional investigations before a fraud can be verified clearly.

• Money Laundering: Money laundering can mean various things depending on who asks, but in a nutshell, it's when criminals hide the true source of funds that they've earned through unsavoury means. This is verifiably converted and sourced from a reliable source. Examples of illegal activity include gambling, corruption, and drug trafficking. The act of changing black money into white money is the only goal of this endeavor. Many legal and financial ramifications stem from the gravity of the crime of money laundering.

• Online Banking Fraud: Telephone banks, mobile banks, and other online banking services have made life much easier for bank consumers. Businesses are offered a choice that is easier, smoother, and more pleasant. Nevertheless, internet banking clients also want security. As the number of transactions conducted online continues to climb, so does the frequency with which fraudsters develop new methods. Users start to question the security of the online banking system after suffering a financial loss due to fraud.

B. Characteristics of the Financial Fraud Data

The multiple characteristics of the financial frauds are discussed below:

• Class imbalance: A common issue with financial fraud datasets is the class imbalance problem, which causes a bias in predictions towards the non-fraud class and poor performance for the fraud class. In the financial sector, detecting fraud is an important yet difficult endeavour. Conventional ML approaches have their work cut out for them when dealing with illicit organizations' clandestine operations and the complicated and sometimes uneven character of transaction data.

• Concept drift: The constant creation, high real-time processing needs, and complicated distributions of streaming data make it subject to concept drift in fields such as healthcare, industrial equipment maintenance, and fraud detection. The notion drift may be broadly classified into two categories:

• Concept drift adaptation: The purpose of concept drift adaptation is to increase the model's adaptability to various input types. It is not the presence or absence of drift that matters with streaming data, but rather the model's passive modification and adaptation to the present data. Specifically, this calls for enhancing the model's generalizability while also reflecting the present facts.

• Concept drift detection: The main objective of concept drift detection which is more proactive is to determine whether the data stream has experienced idea drift. The key objective of drift detection is to find out whether concept drift has occurred at any point in time, as well as the specific time intervals, duration and other relevant information of the drift.

Concept drift identification and adaptation are not opposing. The model provides better representational abilities and tolerance to erroneous detection of drift, which in turn facilitates the adaptive qualities [13]. Besides, early detection saves time and enhances the concept drift capability of the model in streaming data because it avoids useless updates to the model. Due to this, the idea drift control approach is more flexible and effective. It is, therefore, the main concern of this essay that the idea drift is identified.

• Real-time detection requirements: Identity theft and phishing attacks, synthetic fraud, complex money laundering are only some of the elements that make the threat spectrum encountered by banks, fintech companies, and regulatory authorities as dynamic and insidious as never before. Traditional approaches to fraud detection that rely mainly on rule-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

based systems and post-facto investigations are no longer that effective in detecting and preventing threats when they happen [14]. Therefore, Artificial Intelligence (AI) is quickly turning into a disruptive technology, changing the nature of real-time fraud detection in the digital age.

• Privacy regulations and intrusiveness: Privacy regulations protect consumers and influence consumers' privacy awareness. The issues related to privacy intrusion are also strongly associated with the prevention of CCF [15]. A considerable number of scholars have examined the problem of fraud detection, and the algorithms employed are persistently refined to integrate emerging techniques and tactics utilized by fraudsters.

C. Data Mining Techniques used for Financial Fraud Detection

Several new methods have been developed to bolster the development of the six data mining classes that are the foundation for using data mining techniques to detect financial crime.

• Data mining classes: Data mining is a technique for systematically exploring large datasets for meaningful outliers or distortions. The groundwork for this perspective is provided by data mining classes. Problems with detecting financial fraud may be effectively addressed by utilizing six data mining classes.

• Classification: The capacity of a classification model to foretell the categorical class labels of unknown objects is the foundation of its capability to differentiate between things belonging to different classes. Through the process of classification, may learn about a set of commonalities and models that define and differentiate various data kinds. Many other types of fraud may be detected using these classification groups, including healthcare fraud, business fraud, credit card fraud, and vehicle insurance fraud.When considering data mining's role in detecting financial fraud (FFD), it is a crucial strategy.

• Prediction: Prediction models are able to anticipate unknown or missing values, as they are continuous-valued functions. It is necessary for the object's attributes to be continuous, not categorical or discrete valued, in order to make predictions. This characteristic is sometimes known as the expected characteristic. Prediction methods such as logistic model prediction and neural networks are widely utilized.

• Regression: One data mining function is regression, which may forecast a numerical value. Profit, sales, mortgage values, housing expenditures, square footage, temperature, distance, and countless other factors may be easily forecasted using regression algorithms. A house's estimated market worth may be determined using a regression model by considering a variety of criteria, such as its location, square footage, number of bedrooms, and other relevant features. When the goal values are known going into a regression job, the data set is prepared with those values already in it. One common use of the statistical tool known as regression is to describe the connection between two or more independent or dependent variables with discrete values.

III. EMERGING DEEP LEARNING APPROACHES FOR FRAUD DETECTION

The domain of data analysis has been utterly transformed by DL [16], a branch of AI, which allows models to autonomously discover complex patterns inside massive datasets. Contrary to the conventional ML algorithms, which need human intervention in a feature extraction, the DL models are able to discover the features on their own and therefore are most useful in complex and non-linear data such as those in fraud detection [17]. The ability of the DL techniques like CNNs, RNNs and autoencoders to process unstructured data, anomalies, and subtle patterns of fraud renders them invaluable to a modern system of fraud detection. The underlying DL techniques are shown in Figure 2.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

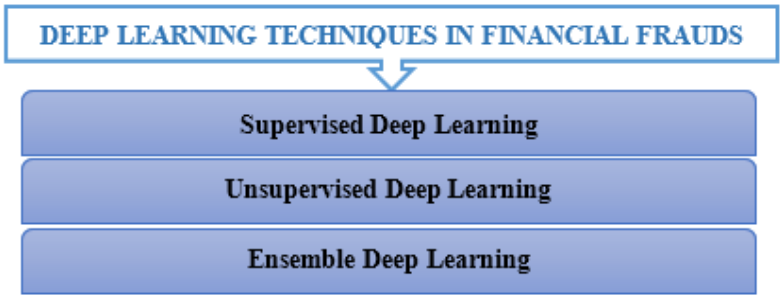

Fig. 2. Deep Learning Methods for Detecting Fraud in Financial Systems

A. Supervised Deep Learning in Detecting Financial Frauds

The financial transactions environment has consequently grown increasingly more complex due to the widening scope of worldwide economic integration and technological innovation of information systems [18]. This complexity complicates the process of detecting and managing financial fraud.

• Convolutional Neural Networks (CNN): CNNs are types of DL models that are specifically devised to handle input data that has a lattice form (i.e., images, time series, or text sequences). CNN first found great success in computer vision, and they have subsequently found widespread usage in voice recognition, NLP, and financial data analysis. In practice, it is not common to apply CNN directly to portfolio optimization, as portfolio optimization usually involves statistics, mathematical optimization methods and traditional ML algorithms. However, CNN can extract valuable information from massive amounts of unstructured data, which can be further integrated into the investment decision-making process, e.g. for enhancing factor analysis or constructing new quantitative strategies.

• Recurrent Neural Network (RNN): The temporal dependency of sequential data may be captured by an RNN, a specific kind of NN architecture. RNN introduces a cyclic structure in the network so that the output of the current moment depends not only on the current input, but is also related to the hidden state of the previous moment, so that past sequential information can be remembered. Recurrent neural networks, with their ability to capture time-series dependencies, have a promising application in portfolio management, especially when dealing with financial data with time-ordered correlations, which can help investors better grasp market dynamics and trends. The architecture of RNN is very flexible, so it can handle sequence inputs of lengthy sequences, and simultaneously consider the backward and forward links among the information at each time point when dealing with sequence data.

• Graph Neural Networks (GNN): Graph Neural Networks (GNNs) are a type of subset of DL models that are optimized to handle graph-based data, or a type of network of connected nodes and edges [19]. Because of their capacity to analyze the network structure of financial transactions and uncover patterns and outliers that traditional rule-based and machine-learning methods might fail to detect, GNNs have become a useful tool to detect fraudulent activity in complex financial systems.

• Long Short-Term Memory: The LSTM neural networks are another type of RNN that have been especially useful in time-series processing, and they have been of interest due to their capability to capture long-term dependencies. LTSTM networks can also be used in portfolio management to forecast future price changes in financial assets e.g. stocks, futures, foreign exchange, etc. to give an investor a guide to future price changes and help them buy and sell and manage risks. By analysing historical price series and related market data, LSTM can also identify patterns and trends in asset price fluctuations, helping to quantify and predict portfolio risk levels such as volatility, maximum retracement, etc.

B. Unsupervised Deep Learning in Detecting Financial Frauds

Unsupervised methods enable learning algorithms to uncover hidden structures within unlabelled datasets. When data is not marked as normal or abnormal, it is unlabelled. The goal of anomaly detection is to spot data events that don't fit the norm. Human expertise is crucial for the discovery or creation of cases based on particular abnormalities when

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

these approaches are used in the aviation sector [20]. Since it would be unrealistic to collect data on every single outlier flight, picking the correct examples to label is crucial for achieving accurate classification with little labelling effort.

• Autoencoder: A popular neural network-based unsupervised learning dimensionality reduction approach is the autoencoder algorithm. Autoencoders [21] and principal component analysis (PCA) [22] both employ neural networks to reduce the dimensionality of data and remove linear limitations from high-dimensional data sets. The encoder and the decoder are the two main parts of an autoencoder, which is designed using a symmetrical network topology. The input data is initially high-dimensional, but the encoder reduces its size. The decoder then reconstructs the original, high-dimensional input data.

• Variational Autoencoder (VAE): The VAE is a deep generative model that combines standard autoencoder architecture with the probabilistic framework of a Bayesian latent space modeling technique. Virtual autoencoders (VAEs) learn to estimate the probability distribution of latent variables, as opposed to traditional autoencoders, which compress input into a single point in latent space. This is a key factor that allows the model to generate new, invisible instances that belong to the same distribution as the original data and detect distortions that are not within this distribution [23]. Financial fraud detection benefits greatly from this capability since it allows the VAE to learn about normal transaction behavior without needing to identify illegal transactions.

C. Ensemble Deep Learning in Detecting Financial Frauds

A strong way to improve classification accuracy is to use ensemble methods. The basic idea of ensemble learning is to improve prediction results by combining several classifiers with different learning processes or training samples. Ensemble learning uses a combination technique or voting system to bring together several supervised or unsupervised classification methods, with the goal of improving the overall performance of the system. In the first step of building an ensemble model, a number of distinct classifiers—usually weak ones—are trained on the training data in order to identify patterns in the data using their own methods.

• Google Net: A CNN design that excels in classification. Created by specialists at Google. The utilization of inception modules, which are blocks housing several parallel convolutional layers with varying filter sizes, is a key component of Google Net. With these modules, the network may simultaneously collect characteristics of varying sizes, improving its representation of the input data. The conception modules were designed to find a middle ground between computational economy and expressive capability. Google Net could be run on computers with a significantly smaller number of parameters compared to standard deep networks due to its utilization of global average pooling, and could be run on significantly more powerful computers. The architecture additionally proposed the idea of introducing auxiliary classifiers at the training layers in the middle, which helped to reduce the vanishing gradient problem and enhance convergence.

• Dense Net: A DL architecture that is attempting to address problems by reusing information flow and features. levels in Dense Net accept data on all levels at all levels below them and contribute to the feature maps of all levels above them. This thick connection not only helps in the vanishing gradient problem but also helps in reusing features and the flow of information that is very efficient within the network. The architecture of the Dense Net has proven to be a competitive model in terms of reduced parameters when compared to other DL architectures, it also encourages the flow of parameters and also helps to propagate the features. Models trained with Dense Net tend to be smaller, more accurate, and easier to train because of the network's dense connections.

• VGG: The Visual Geometry Group entered the 2014 ImageNet Large Scale Visual Recognition Challenge with a CNN architecture. The consistent architecture and simplicity of VGG are its defining features. Throughout its creation, VGG kept to a consistent structure, in contrast to other modern models that incorporated complicated topologies and different sorts of layers. The network was made deeper by stacking several convolutional layers with modest filter sizes and mainly used 3x3 convolutional filters. VGG architecture variations. VGG architecture has many variants, with VGG16 and VGG19 as the most important ones, indicating the number of weight layers in each version. VGG16 and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

VGG19 have 16 and 19 layers of weights respectively. Subsequent designs that sought to balance depth, complexity, and computational efficiency of DL models in image recognition can be viewed as VGG influence.

• Res Net: The concept of residual learning was introduced by ResNet, a DL system. Microsoft Research Associates created it. The usage of residual blocks is ResNet's main innovation. When training extremely deep networks, conventional DNNs also fail due to vanishing gradients and a decline in accuracy. Residual learning is one approach to these issues; it allows the network to bypass some levels in forward and backward propagation by adding shortcut connections, sometimes called skip connections. The two information flow channels in a residual block are one with standard convolutional layers and the other with a direct shortcut, where the original input is added to the processed output. The former is a sort of shortcut link. This architecture permits the learning of very deep networks by evading the vanishing gradient problem and transferring the gradient directly down the shortcut links. The total number of layers in a ResNet-based architecture determines its depth; for instance, ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 are all examples of depths. Object recognition, classification, and image segmentation are just a few computer vision tasks that have shown remarkable results when trained on these systems.

IV. BENCHMARKING DATASETS IN FINANCIAL FRAUD DETECTION

The datasets used for training and assessment have a significant impact on the models' performance and practical applicability due to their quality, attributes, and representativeness. The overall analysis revealed that there are a number of datasets that are widely applied in fraud detection studies and have their unique features that are pertinent to various types of fraud and detection situations [24]. This section explains the different datasets used in Financial Fraud Detection:

A. Credit Card Fraud Detection Datasets

The financial sector has undergone a fundamental shift due to the proliferation of online payment systems, which have made transactions much simpler and more efficient. Credit card has become the foundation of the current business with this revolution as the payment method is easy to use in purchasing goods and services. However, a major and extremely costly problem in the financial sector, credit card theft has increased in tandem with the proliferation of these digital financial solutions.

• Credit Card Fraud Detection Datasets: Credit card fraud detection research relies on the dataset maintained by the Machine Learning Group at Université Libre de Bruxelles. This dataset includes anonymized credit card transactions made by European consumers in September 2013 and has been referenced in fifteen studies [25]. The number of fraudulent transactions out of 284,807 was 492 which was a fraud rate of 0.172%. The majority of characteristics were turned into secret with the help of principal component analysis (PCA) with the exception of Time and Amount, which would remain original features.

• German Credit: This credit risk classification dataset was created by Professor Hofmann for the UCI Machine Learning Repository [26]. It has 1,000 instances, and each one is represented by 20 variables that show personal details and credit information.

• Australian Credit Approval: The UCI Machine Learning Repository contains a dataset including 690 credit card application instances with 14 associated characteristics [27]. Studies have used it.

• Default of Credit Card Clients: Credit card clients in Taiwan who have failed on their payments are the subject of this dataset from the UCI ML repository [26]. It spans the months of April through September 2005 and contains 30,000 occurrences with 24 characteristics pertaining to payment and credit history.

B. Financial Statement Fraud Datasets

Financial statement fraud is a white-collar crime that is usually committed by people in management to make the company look better financially. Motives for fraud can range from short-term gains (such as performance-based

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

reward) to long-term gains (such as manipulating potential investors or improving the company's reputation) or even just buying time until financial errors and losses can be adequately rectified.

• China Stock Market and Accounting Research (CSMAR): This extensive database covers the years 1998–2016 and includes financial statements of listed firms in China's stock marketplaces [28]. Research using this dataset comprises 35,574 samples representing 337 instances of fraud every year.

• CompStat: There is economic and financial data about US and Canadian businesses in this database. Included in the dataset are details on 228 different organizations [29], of which 50% had instances of bank account fraud. Various studies have made use of it.

C. Synthetic Datasets

The term "synthetic dataset" describes data sets that have been "created" in a lab using computational methods. In fields where it is difficult to get data from the actual world, such as surgery, they come in quite handy. To get beyond problems like noise and camera distortion that are common in real-world datasets, researchers use synthetic datasets to test point cloud registration algorithms.

• PaySim Mobile Money Simulator: An artificial dataset was constructed using data from a mobile money provider in an African country. In it are 6,362,620 samples with a total of 8,213 instances of fraudulent transactions.

• BankSim Payment Simulator: This synthetic dataset contains 594,643 transactions derived from a subset of a Spanish bank's data. Approximately 1,2% (7,200) of these transactions were marked as fraudulent [30].

D. Other Specialized Datasets

The financial scams are detected using a variety of different databases. What follows is an in-depth discussion of these datasets:

• Insurance Company Benchmark (COIL2000): Information on insurance business clients, including demographics, product use, and insurance claims history, is included in this dataset [31]. Studies involving insurance fraud detection have made use of its 9822 occurrences and 86 characteristics.

• Bitcoin Network Transactional Metadata: The dataset includes information about Bitcoin transactions and flows from 2011 to 2013, with 30,000 occurrences and 11 characteristics linked to the cryptocurrency [32]. Its usage includes the examination of suspicious Bitcoin transactions.

Financial Fraud Detection makes use of the datasets summarized in Table I

TABLE I. COMMONLY USED DATASETS IN FINANCIAL FRAUD DETECTION

Category	Dataset Name	Source	Size / Instances	Key Characteristics	Application Area
Credit Card Fraud	CCF Detection Dataset	Université libre de Bruxelles (ULB)	284,807 transactions (492 frauds)	Highly imbalanced dataset; PCA-transformed features; original features include Time and Amount	Credit card fraud detection
	German Credit Dataset	UCI ML Repository	1,000 instances, 20 attributes	Credit risk attributes describing personal and financial information	Credit risk and fraud analysis
	Australian Credit Approval Dataset	UCI ML Repository	690 instances, 14 attributes	Credit application-related attributes	Credit approval and fraud detection
	Default of Credit Card	UCI ML Repository	30,000 instances,	Credit data and payment history from Taiwan customers	Default prediction and fraud detection

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

	Clients		24 attributes		
Financial Statement Fraud	CSMAR	CSMAR Database	35,574 samples (337 fraud cases)	Financial statements of Chinese listed firms (1998–2016)	Financial statement fraud detection
	Compustat	Compustat Database	228 companies	Financial and economic data with labeled fraud cases	Corporate and accounting fraud detection
Synthetic Datasets	PaySim Mobile Money Simulator	Simulated mobile money data	6,362,620 transactions (8,213 frauds)	Synthetic dataset simulating real mobile money transactions	Transaction and mobile payment fraud
	BankSim Payment Simulator	Simulated bank transactions	594,643 transactions (~1.2% fraud)	Synthetic payment data based on Spanish bank transactions	Banking transaction fraud detection
Other Specialized Datasets	Insurance Company Benchmark (COIL2000)	UCI ML Repository	9,822 instances, 86 attributes	Customer, product usage, and sociodemographic data	Insurance fraud detection
	Bitcoin Network Transactional Metadata	Bitcoin Transaction Dataset	30,000 instances, 11 attributes	Cryptocurrency transaction flow and metadata	Cryptocurrency fraud and anomaly detection

V. LITERATURE REVIEW

This paper was informed and strengthened by a literature review focusing on techniques and challenges in fraud detection within financial systems. Table II presents the summary of the existing literature showing the dataset description, model and method, key features and contributions.

Tang (2025) researches how to use the Graph AGE model to identify fraudulent financial transactions in real-time social networks. The dataset, sourced from various business periods of enterprises, constructs a fully connected directed dynamic graph of a social network. Nodes represent registered users, while edges denote emergency contact relationships between users and include time and type attributes. Their objective is to improve the accuracy of financial fraud detection by identifying fraudulent users within these graph structures. In the feature engineering phase, they extracted complex features such as node in-degrees and out-degrees, edge types, and time information [33].

Sargam and Kalapala (2025) this study assume an integrated federated anomaly detection system with cloud coordination on the Amazon Web Services (AWS) to facilitate safe and cooperative fraud detection among several healthcare organizations. In the mechanism, the unsupervised deep auto encoders are used to detect anomalies and Federated Averaging (FedAvg) is used to facilitate decentralized model training without sacrificing raw data sharing. Differential privacy controls are implemented in order to preserve sensitive data during the learning process. The system operates within the AWS framework, where the calculations are performed using EC2 instances, the model is hosted using SageMaker, and data is stored in Amazon S3, where it is safely stored. Real insurance claim data were experimented with artificial instances of fraud that mimic realistic patterns of fraud [34].

Panda, Ojha and Dhal (2024) propose a new machine learning architecture that combines ISSO and PKRR to improve fraud detection. The proposed model enhances the detection rates and reduces the false positives to a considerable

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

extent by combining the derivation of the ISSO with the high levels of the regression and partitioning of the PKRR. This combination removes the current shortcomings and provides a flexible, scalable and viable solution to protecting the financial systems against the emerging fraud threats [35].

Kesharwani and Shukla (2024). The advanced framework of detection of fraud in transactions is created with the help of FFDM, employing GNNs to recognize fraud cases in transactions. The traditional fraud detection systems tend to encounter the problem of dynamism and complexity of fraud. In this study, a novel framework of fraud detection based on exploiting GNN to address such issues is presented. Trying to solve the highly complex patterns and interactions of monetary transactions, they have created a complementary model that integrates the most appropriate aspects of graph-based learning with DNN [36].

Somkunwar et al. (2023) detecting financial fraud within banking infrastructure and internet transacting records is a multifaceted issue that needs strong methodologies. In financial network analysis, a technique known as graph mining is utilized to identify connections and patterns through data exploration; it is also applicable in other fields of data exploration like data visualization. To determine the suspicious monetary transactions, the Antireform subgraph was built on the basis of the Binford Law. False positives on the other hand have been reported. One of the possible solutions to this problem is the use of the Antireform subgraph together with an unsupervised MLA. In case of concern regarding the security of the financial networks, require this new fraud detection system [37].

Biswas et al. (2022) Using data mining, banks can better spot customers who may be trying to defraud them when they apply for loans. Banks use data mining tools like support vector machines (SVMs), logistic regression (logistics), decision trees (DTs), to spot fraudulent activity. But before building a model, each of these tools needs a data balancing method. Without data balance, autoencoder is one method that has been proposed for fraud detection. Using a method for secondary thematic data analysis, this study investigates a number of aspects connected to the topic. Risk in the banking industry has been successfully mitigated by the implementation of the suggested model for cybercrime detection and an automated fraud regulating system enabled by AI [38].

TABLE II. SUMMARY OF LITERATURE STUDIES ON FRAUD DETECTION WITHIN FINANCIAL SYSTEMS

Author(s) & Year	Application Domain	Dataset Description	Methodology / Model	Key Features Used	Main Contribution
Tang (2025)	Financial fraud in social networks	Enterprise user data across multiple business periods	GraphSAGE-based Graph Neural Network	Node in-degree, node out-degree, edge type, temporal features	Improved fraud detection accuracy by modeling dynamic user relationships in fully connected directed graphs
Sargam & Kalapala (2025)	Healthcare insurance fraud	Real insurance claims with synthetic fraud samples	Federated anomaly detection using autoencoders and FedAvg on AWS	Unsupervised learning, federated training, differential privacy	Enabled privacy-preserving collaborative fraud detection using cloud-based federated learning infrastructure
Panda, Ojha & Dhal (2024)	Financial transaction fraud	Financial transaction datasets	Hybrid ISSO–PKRR machine learning framework	Optimization-based feature selection, robust regression, data partitioning	Increased fraud detection rate and reduced false positives with a scalable and adaptive framework
Kesharwani & Shukla	Financial transaction	Transaction-level	Graph Neural Network–	Graph-based features, deep	Effectively captured complex transactional

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

(2024)	networks	financial data	based FFDM	neural network representations	patterns, overcoming limitations of traditional fraud detection systems
Somkunwar et al. (2023)	Banking and digital ledger fraud	Banking transaction and ledger datasets	Graph mining + Unsupervised Machine Learning Algorithm	AntiBenford subgraph, unsupervised learning	Reduced false positives of Benford's Law and strengthened fraud detection in financial networks
Biswas et al. (2022)	Banking sector fraud	Credit and banking transaction data	Data mining and AI models (Autoencoder, SVM, LR, DT, NN)	Autoencoder without data balancing, thematic data analysis	Demonstrated effectiveness of autoencoders and AI-enabled fraud control systems in reducing banking risks

VI. CONCLUSION AND FUTURE WORK

Financial fraud identification is a serious field where ever-changing fraudulent schemes require high-level and dynamic identification processes. The complexity and severity of fraud in today's banking, insurance, internet transactions, and money laundering industries are on the rise, and this paper gives a comprehensive review of financial fraud detection in these systems. It also analyzed the nature of financial fraud, the nature of the challenges that face fraud data, including the skew in classes, concept drift, the need to detect fraud in real time, and privacy limitations, and how data mining and DL methods can be used to solve these problems. The discussion has shown that although other machine learning methods have a background structure, new DL models, such as CNNs, RNNs, GNNs, autoencoders, and ensemble models, have better potential to capture intricate patterns, temporal relations, and relational arrangements in financial data. Moreover, research on benchmark data sets supports the necessity of having a variety of high-quality data in assembling generalizable fraud detection systems, whether it be of real financial statement and credit card data or of artificial simulators.

The future research can be guided by developing adaptive privacy-preserving models of fraud detection that are capable of processing streaming data and concept drift more efficiently. Federated learning, graph-based DL, and explainable AI should be integrated to make future financial fraud detection systems all three times the scalability, transparency, and regulatory compliance.

REFERENCES

[1] S. B. Shah, "Advancing Financial Security with Scalable AI: Explainable Machine Learning Models for Transaction Fraud Detection," in 2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), IEEE, Apr. 2025, pp. 1–7. doi: 10.1109/ICDCECE65353.2025.11034838.

[2] A. Parupalli, "The Evolution of Financial Decision Support Systems : From BI Dashboards to Predictive Analytics," KOS J. Bus. Manag., vol. 1, no. 1, pp. 1–8, 2023.

[3] F. Xu and R. Zhang, "Explainable Domain Adaptation Learning Framework for Credit Scoring in Internet Finance Through Adversarial Transfer Learning and Ensemble Fusion Model," Mathematics, vol. 13, no. 7, 2025, doi: 10.3390/math13071045.

[4] V. Verma, "Deep Learning-Based Fraud Detection in Financial Transactions : A Case Study Using Real-Time Data Streams," vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[5] T. Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," Int. J. Sci. Res. Arch., vol. 15, no. 3, pp. 618–627, Jun. 2025, doi: 10.30574/ijsra.2025.15.3.1767.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

[6] K. M. R. Seetharaman, "Digital Transformation in Retail Sales: Analyzing the Impact of Omni-Channel Strategies on Customer Engagement," J. Glob. Res. Math. Arch., vol. 10, no. 12, 2023, doi: 10.5281/zenodo. 15280578.
[7] S. Garg, "Predictive Analytics and Auto Remediation using Artificial Intelligence and Machine learning in Cloud Computing Operations," Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 7, no. 2, 2019, doi: 10.5281/zenodo.15362327.
[8] S. J. Wawge, "A Survey on the Identification of Credit Card Fraud Using Machine Learning with Precision, Performance, and Challenges," Int. J. Innov. Sci. Res. Technol., vol. 10, no. 4, pp. 3345–3352, May 2025, doi: 10.38124/ijisrt/25apr1813.
[9] R. P. Mahajan, "Optimizing Pneumonia Identification in Chest X-Rays Using Deep Learning Pre-Trained Architecture for Image Reconstruction in Medical Imaging," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 1, pp. 52–63, Apr. 2025, doi: 10.48175/IJARSCT-24808.
[10] R. P. Mahajan and N. Jain, "Optimizing CT Image Quality through AI-based Reconstruction and Deep Learning Models for Enhanced Diagnostic Accuracy," in 2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2025, pp. 1–7. doi: 10.1109/ICDCECE65353.2025.11035138.
[11] D. Patel, "Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model," Int. J. Curr. Eng. Technol., vol. 13, no. 06, pp. 576–583, 2023, doi: 10.14741/ijcet/v.13.6.10.
[12] Y. Macha and S. K. Pulichikkunnu, "An Explainable AI System for Fraud Identification in Insurance Claims via Machine-Learning Methods," Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 3, pp. 1391–1400, Jul. 2023, doi: 10.48175/IJARSCT-11978X.
[13] L. Hu, Y. Lu, and Y. Feng, "Concept Drift Detection Based on Deep Neural Networks and Autoencoders," Appl. Sci., vol. 15, no. 6, 2025, doi: 10.3390/app15063056.
[14] P. S. Raibagi, "Artificial Intelligence and Financial Fraud Detection : A Review Paper," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 7, pp. 1–9, 2025, doi: 10.48175/IJARSCT-26801.
[15] L. Gabudeanu, I. Brici, M. Codruta, I.-C. Mihai, and M. Scheau, "Privacy Intrusiveness in Financial-Banking Fraud Detection," Risks, vol. 9, p. 104, 2021, doi: 10.3390/risks9060104.
[16] J. Kachhia, R. Natharani, and K. George, "Deep Learning Enhanced BCI Technology for 3D Printing," in 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, Oct. 2020, pp. 0125–0130. doi: 10.1109/UEMCON51285.2020.9298124.
[17] M. Blessing, "Enhancing Fraud Detection with Deep Learning: An In-Depth Analysis," 2024.
[18] R. Patel, "Automated Threat Detection and Risk Mitigation for ICS (Industrial Control Systems) Employing Deep Learning in Cybersecurity Defence," vol. 13, no. 6, pp. 584–591, 2023.
[19] V. Pal and S. K. Chintagunta, "Transformer-Based Graph Neural Networks for Real-Time Fraud Detection in Blockchain Networks," Int. J. Adv. Res. Sci. Commun. Technol., pp. 1401–1411, Jul. 2023, doi: 10.48175/IJARSCT-11978Y.
[20] P. B. Patel, "Energy Consumption Forecasting and Optimization in Smart HVAC Systems Using Deep Learning," Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 3, pp. 780–788, 2024, doi: 10.48175/IJARSCT-18991.
[21] U. A. Korat and A. Alimohammad, "A Reconfigurable Hardware Architecture for Principal Component Analysis," Circuits, Syst. Signal Process., vol. 38, no. 5, pp. 2097–2113, May 2019, doi: 10.1007/s00034-018-0953-y.
[22] H. Du, L. Lv, A. Guo, and H. Wang, "AutoEncoder and LightGBM for Credit Card Fraud Detection Problems," Symmetry (Basel)., vol. 15, no. 4, p. 870, Apr. 2023, doi: 10.3390/sym15040870.
[23] S. Obushnyi, D. Virovets, A. Ramskyi, and M. Zhytar, "Variational Autoencoders for Detecting Anomalous and Fraudulent Transactions in Financial Systems," Digit. Econ. Concepts Technol., pp. 110–118, 2025.
[24] A. R. Bilipelli, "Application of AI and Data Analysis for Classification of Student Success in Large-Scale Educational Dataset," Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 6, pp. 428–441, Nov. 2024, doi: 10.48175/IJARSCT-22564.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

[25] M. R. Baker, Z. N. Mahmood, and E. H. Shaker, "Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions," Rev. d'Intelligence Artif., vol. 36, no. 4, pp. 509–518, Aug. 2022, doi: 10.18280/ria. 360401.
[26] 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, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.
[27] A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 18–25, 2018, doi: 10.14569/IJACSA.2018.090103.
[28] Y. Chen and Z. Wu, "Financial Fraud Detection of Listed Companies in China: A Machine Learning Approach," Sustainability, vol. 15, no. 1, 2023, doi: 10.3390/su15010105.
[29] I. Dutta, S. Dutta, and B. Raahemi, "Detecting financial restatements using data mining techniques," Expert Syst. Appl., vol. 90, pp. 374–393, Dec. 2017, doi: 10.1016/j.eswa.2017.08.030.
[30] M. Seera, C. P. Lim, A. Kumar, L. Dhamotharan, and K. H. Tan, "An intelligent payment card fraud detection system," Ann. Oper. Res., 2024, doi: 10.1007/s10479-021-04149-2.
[31] D. Huang, D. Mu, L. Yang, and X. Cai, "CoDetect: Financial Fraud Detection with Anomaly Feature Detection," IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2816564.
[32] T. Ashfaq et al., "A Machine Learning and Blockchain-Based Efficient Fraud Detection Mechanism," Sensors, vol. 22, no. 19, p. 7162, Sep. 2022, doi: 10.3390/s22197162.
[33] J. Tang, "Application of GraphSAGE in Financial Fraud Detection within Dynamic Social Networks," in 2025 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE), 2025, pp. 989–993. doi: 10.1109/EDPEE65754.2025.00179.
[34] G. S. Sargam and R. Kalapala, "AI-Driven Claim Fraud Detection in Health Insurance Using Federated Anomaly Detection Networks with Cloud Computing on AWS for Privacy-Preserving Financial Security," in 2025 Third International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV), 2025, pp. 1–6. doi: 10.1109/ICPEEV67897.2025.11291290.
[35] G. Panda, R. K. Ojha, and S. K. Dhal, "A Machine Learning-Based Approach to Improve Fraud Detection using ISSO and PKRR," in 2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), 2024, pp. 1–6. doi: 10.1109/IC-SIT63503.2024.10862834.
[36] A. Kesharwani and P. Shukla, "FFDM − GNN:A Financial Fraud Detection Model using Graph Neural Network," in 2024 International Conference on Computing, Sciences and Communications (ICCSC), 2024, pp. 1–6. doi: 10.1109/ICCSC62048.2024.10830438.
[37] R. K. Somkunwar, A. Pimpalkar, K. M. Katakdound, A. S. Bhide, S. P. Chinchalkar, and Y. M. Patil, "A Fraud Detection System in Financial Networks Using AntiBenford Subgraphs and Machine Learning Algorithms," in 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE), 2023, pp. 1–6. doi: 10.1109/AIKIIE60097.2023.10390325.
[38] A. Biswas, R. S. Deol, B. K. Jha, G. Jakka, M. R. Suguna, and B. I. Thomson, "Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector," in 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), 2022, pp. 809–814. doi: 10.1109/ICOSEC54921.2022.9951931.