applied sciences

MDPI

Review

Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review

Abdulalem Ali [1,*], Shukor Abd Razak [1,2,*], Siti Hajar Othman [1], Taiseer Abdalla Elfadil Eisa [3], Arafat Al-Dhaqm [1,*], Maged Nasser [4], Tusneem Elhassan [1], Hashim Elshafie [5] and Abdu Saif [6]

1 School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai 81310, Malaysia
2 Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu 21300, Malaysia
3 Department of Information Systems-Girls Section, King Khalid University, Mahayil 62529, Saudi Arabia
4 School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Malaysia
5 College of Computer Science, King Khalid University, Abha 61421, Saudi Arabia
6 Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
* Correspondence: amsabdulalem2@graduate.utm.my (A.A.); shukorar@utm.my or shukorrazak@unisza.edu.my (S.A.R.); mrarafat1@utm.my (A.A.-D.)

Abstract: Financial fraud, considered as deceptive tactics for gaining financial benefits, has recently become a widespread menace in companies and organizations. Conventional techniques such as manual verifications and inspections are imprecise, costly, and time consuming for identifying such fraudulent activities. With the advent of artificial intelligence, machine-learning-based approaches can be used intelligently to detect fraudulent transactions by analyzing a large number of financial data. Therefore, this paper attempts to present a systematic literature review (SLR) that systematically reviews and synthesizes the existing literature on machine learning (ML)-based fraud detection. Particularly, the review employed the Kitchenham approach, which uses well-defined protocols to extract and synthesize the relevant articles; it then report the obtained results. Based on the specified search strategies from popular electronic database libraries, several studies have been gathered. After inclusion/exclusion criteria, 93 articles were chosen, synthesized, and analyzed. The review summarizes popular ML techniques used for fraud detection, the most popular fraud type, and evaluation metrics. The reviewed articles showed that support vector machine (SVM) and artificial neural network (ANN) are popular ML algorithms used for fraud detection, and credit card fraud is the most popular fraud type addressed using ML techniques. The paper finally presents main issues, gaps, and limitations in financial fraud detection areas and suggests possible areas for future research.

Keywords: financial fraud; fraud detection; machine learning; data mining; systematic literature review; Kitchenham approach

Citation: Ali, A.; Abd Razak, S.; Othman, S.H.; Eisa, T.A.E.; Al-Dhaqm, A.; Nasser, M.; Elhassan, T.; Elshafie, H.; Saif, A. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Appl. Sci.* **2022**, *12*, 9637. https://doi.org/10.3390/app12199637

Academic Editors: Luis Javier Garcia Villalba, Rafael T. de Sousa, Jr., Robson de Oliveira Albuquerque and Ana Lucila Sandoval Orozco

Received: 23 August 2022
Accepted: 20 September 2022
Published: 26 September 2022

1. Introduction

Financial fraud is the act of gaining financial benefits by using illegal and fraudulent methods [1,2]. Financial fraud can be committed in different areas, such as insurance, banking, taxation, and corporate sectors [3]. Recently, financial transaction fraud [4], money laundering, and other types of financial fraud [5] have become an increasing challenge among companies and industries [4]. Despite several efforts to reduce financial fraudulent activities, its persistence affects the economy and society adversely, as large amounts of money are lost to fraud every day [6]. Several fraud detection approaches were introduced many years ago [1]. Most traditional methods are manual, and this is not only time consuming, costly, and imprecise but also impractical [7]. More studies are conducted to reduce losses resulting from fraudulent activities, but they are not efficient [5]. With the advancement of the artificial intelligence (AI) approach, machine learning and data mining have been utilized to detect fraudulent activities in the financial sector [8,9]. Both

unsupervised and supervised methods were employed to predict fraud activities [4,10]. Classification methods have been the most popular method for detecting financial fraudulent transactions. In this scenario, the first stage of model training uses a dataset with class labels and feature vectors. The trained model is then used to classify test samples in the next step [1,2,5].

Thus, this study attempts to identify machine-learning-based techniques employed for financial transaction fraud and to analyse gaps to discover research trends in this area. Recently, some reviews have been conducted to detect fraudulent financial activities [11–13]. For instance, Delamaire et al. [11] conducted a review on different categories of fraudulent activities on credit cards, which include bankruptcy and counterfeit frauds, and suggested proper approaches to address them. Similarly, Zhang and Zhou [12] investigated ML methods for fraud transactions, which include the stock market and other fraud detection processes in financial sectors. Raj and Portia. [13] explored several ML approaches used for credit card fraud detection. Phua et al. [14] conducted a comprehensive survey to explore data mining and machine learning techniques to detect frauds in various aspects, including credit card fraud, insurance fraud, and telecoms subscription fraud.

Recently, there has been a significant increase in fraud activities in health sectors [15]. Abdallah et al. [16] introduced a review to investigate different approaches for uncovering fraudulent activities in the health care domain based on statistical approaches. Popat and Chaudhary [17] presented an extensive review work on credit card fraud detection. The authors provide a detailed analysis of various ML classification methods with their methodology and challenges. Ryman-Tubb et al. [6] reviewed several state-of-the-art methods for detecting payment card fraudulent activities using transactional volumes. The study showed that only eight approaches have a practical implication to be used in the industry. A study by Albashrawi and Lowell [3] analyzed several studies for one decade covering fraud detection in financial sectors using data mining techniques. However, this was not exhaustive and comprehensive enough as they ignored the method of evaluations and the pros and cons of data mining techniques, among others.

Despite several existing reviews in the field, however, most studies particularly focused on specific areas of finance, such as detecting credit card fraudulent activities [18], fraud in online banking [19], fraud in bank credit administration [20], and fraud in payment cards [21]. Hence, there is a need of a study that encompasses all popular areas of financial fraud activities to fill the gap in this aspect. More recently, a study was published to review fraud-detection methods in financial records [2]. The authors integrated the prior multi-disciplinary literature on financial statement fraud. However, there are several differences between their work and our review. First, their primary objective is to integrate research from several fields, including information systems, analytics, and accounting. On the other hand, we aim to identify financial fraud transactions based on machine learning methods and to discover datasets applied in the ML-based financial fraud detection. Furthermore, we considered conference articles in our study while they did not. This study reviews existing machine learning (ML)-based methods applied for financial transaction fraud detection. Furthermore, the SLR can guide researchers in their choice of applying ML-based financial transaction fraud-detection methods along with the datasets to be used for predicting fraudulent activities in financial transactions.

The rest of this paper is organized as follows: Section 2 describes the research methodology, including the search criteria, study selection, data extraction, and quality evaluation. The SLR findings and the responses to the study questions are presented in Section 3. The discussion and possible challenges that undermined the validity of this review are addressed in Sections 4 and 5, respectively. Finally, we provide a conclusion of the study in Section 6.

2. Research Methods

In this paper, an SLR approach is used, which is a detailed approach for gathering and analyzing all studies that focused on specific research questions [22]. It is used to identify

and combine information that focuses on particular issues to lessen biases [17,22], provide a review with high-quality evidence, and inspect the path of reviewers' judgments and conclusions [22]. This SLR study is based on the study in [23], which covers three main stages: review planning, conducting the review, and reporting the review. The main stages of SLR are illustrated in Figure 1.

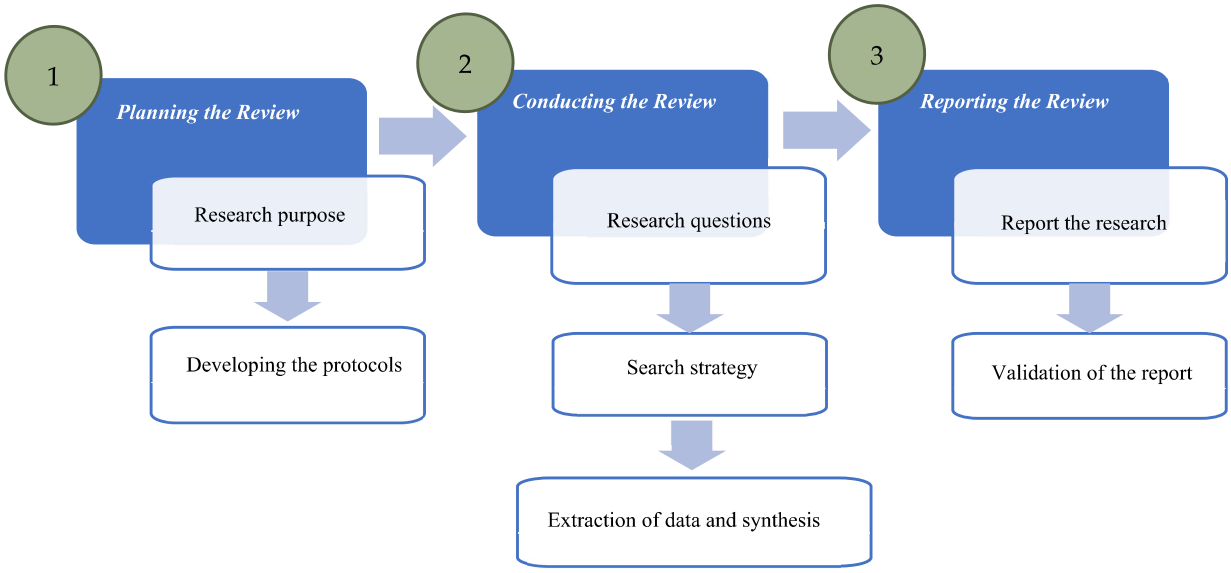

Figure 1. Stages of the SLR.

2.1. Review Planning

The planning stage presents the preparatory and the development processes of the SLR, which involves identifying the research goal and developing the review protocol [2,24]. To obtain more relevant papers, an automatic search was conducted on the major relevant digital databases [25,26]. Other similar databases were not considered as the index data from primary sources. These libraries were considered due to their popularity and being a rich source of articles relevant to the research questions for this study. To obtain comprehensive and up-to-date relevant articles for the SLR, we consider 2010 to 2021 for the review.

2.2. Conducting the Review

After planning the review, the next stage is conducting the review. In this step, the main review process is performed, which involves identifying the research questions (RQs) of the SLR, in which the main issues to be discussed and analyzed in the review are identified. In this stage, the search strategy selection procedure and data extraction and synthesis are presented. These can be explained in the following subsections:

2.2.1. Research Questions

Research questions are first identified in this SLR to identify issues that are addressed and analyzed. The research questions are crucial for identifying the primary studies in the review. Formulating the research question generally forms the main aspect of the SLR. Table 1 presents the main RQs that are used in this study.

The first question helps identify popular financial fraud categories addressed using ML methods. The second question attempts to identify common ML approaches for detecting fraudulent financial activities. The third and fourth questions help identify the performance evaluation metrics used for ML-based financial fraud detection and the research gaps, trends, and future direction in this field.

Table 1. Research Questions.

S/N	RQ	Motivation
1	What popular financial frauds that are addressed based on the ML approaches?	Identify the popular types of financial frauds that are detected based on the ML methods.
2	What popular ML-based approaches employed for financial fraud detection?	Identify the popular categories of the ML methods used for financial fraud detection.
3	What are the evaluation metrics employed to detect financial frauds?	Identify the evaluation metrics used for financial fraud detection.
4	What are the research gaps, trends, and future directions of the research area?	Identify the research gaps, the trends, and future directions in fraud detection research in online transactions.

2.2.2. Search Strategy

To identify the most relevant papers in ML-based financial fraud detection, the authors designed some search terms relevant to the RQs of this study, which involve using Boolean terms, such as "OR" or "AND", to combine search terms that are relevant with the RQs of this SLR. The search terms used in this SLR include the following: "financial fraud" AND "financial transaction" AND "machine learning" OR "artificial intelligence". We searched the above search terms in different popular databases including IEEE Xplore, ACM Digital Library, Web of Science, ScienceDirect, and Scopus. The search terms are modified and converted into appropriate input queries for each digital library search.

2.2.3. Study Selection Criteria

After applying the search terms in the above digital libraries, a total number of 287 papers were discovered from all search databases in which 63 duplicate papers were discovered and filtered from the explored papers. After filtering duplicates, we continued with the selection process using the 124 articles that remained. Authors design inclusion and exclusion criteria in the searching process to identify the most relevant papers. Authors screen these studies following the requirements of the quality assessment standards in order to guarantee the quality of the chosen papers as well. We employ the cross-checking method to determine whether the selected papers match these requirements in order to guarantee the credibility of the results. After applting all the above criteria and the step of quality assessment criteria, 93 studies were finally obtained, which are related to the research questions. Tables 2 and 3 show the inclusion criteria and quality assessment respectively.

Table 2. Exclusion and inclusion criteria.

S/N	Exclusion	Inclusion
1	Articles that do not focus on financial fraudulent transactions.	
2	Articles that are in the form of abstracts, short papers, posters, and book chapters.	The articles that are conducted from 2010 to 2021.
3	Articles that do not pertain to the use of ML/data mining methods.	Articles that focus on financial fraud detection and applied ML methods
4	Studies that do not mention their performance evaluation metrics	A peer-reviewed research article.
5	Studies that were not published in the English language.	Studies were conducted in English only.

Table 3. Quality assessment.

ID	Quality Assessment
1	Is the purpose of the study clear?
2	Are the techniques clearly stated and explained?
3	Are the proposed techniques clearly presented and implemented?
4	Is the experimental procedure clearly described?
5	Does the study make contributions to the SLR?
6	Are empirical experiments clearly stated?
7	Are the performance measures clearly stated?
8	Are the conclusion and future direction clearly stated?

2.3. Data Extraction and Synthesis

The data extraction process involves designing forms to extract information from the selected papers for the SLR [2]. Considering the information in the data extraction form, the specified research questions for the SLR can be answered. The information extracted in the data extraction stage is shown in Table 4.

Table 4. Data extraction form.

Search Method	Information Extracted	Purpose of the Extraction
Manual Search	The category of financial fraud addressed in the study	RQ1
	The technique used for the fraud detection	RQ2
	The objective of the study	RQ1, RQ2
	The evaluation metrics used to address which technique	RQ3
	Future direction, trends, and gaps in the study	RQ4
	Conclusion of the study	RQ1, RQ2, and RQ4
Automatic Search	Title of the study Publication year Names of the author Publication type (conference proceeding or journal article) The conference or journal names	Study Description and Meta-Analysis

In the first column of Table 4, the search strategy used to gather the information for the data extraction including automatic and manual extraction is specified. The second and third columns are the category of the information extracted and the purpose of extracting the information, respectively. This information is analyzed to make it easier for grouping similar studies together in terms of the type of fraudulent activity addressed, the techniques used for the fraud detection, and the evaluation metrics used for the validation method as well as the research gaps and future direction.

3. Search Results and Meta-Analysis

This section presents the search results obtained from the second stage of the review process, which involves selecting the relevant studies to be considered in this SLR study. We first present the description of the reviewed studies in this SLR and then later answer each of the research questions specified in the section.

3.1. Description of Studies

The number of articles relating to financial fraud detection using ML approaches from 2010 to 2021 is shown in Figure 2, which provides a chronological summary of the published articles used in this review. The graph illustrates how research in this field has shown a growing trend in recent years, particularly since 2013 when the number of articles published began to rise significantly. Most articles used in this review were released after 2013. Within the study period, 2016 experienced the highest number of articles (11) published in this area, followed by 10 papers in 2018 and 2017. It can be observed that there was a lower rate of relevant publications in 2013 as only four papers were reviewed in that year. Figure 3 shows various journal names and the number of relevant articles that were used in this review.

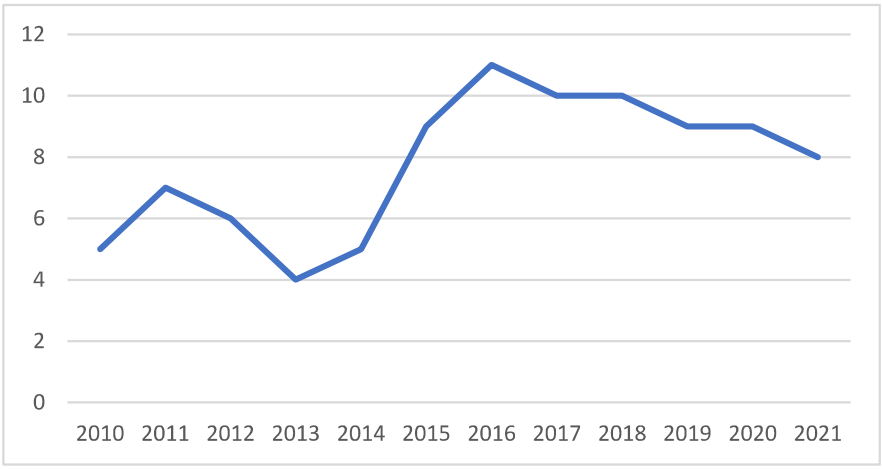

Figure 2. Summary of articles in a year.

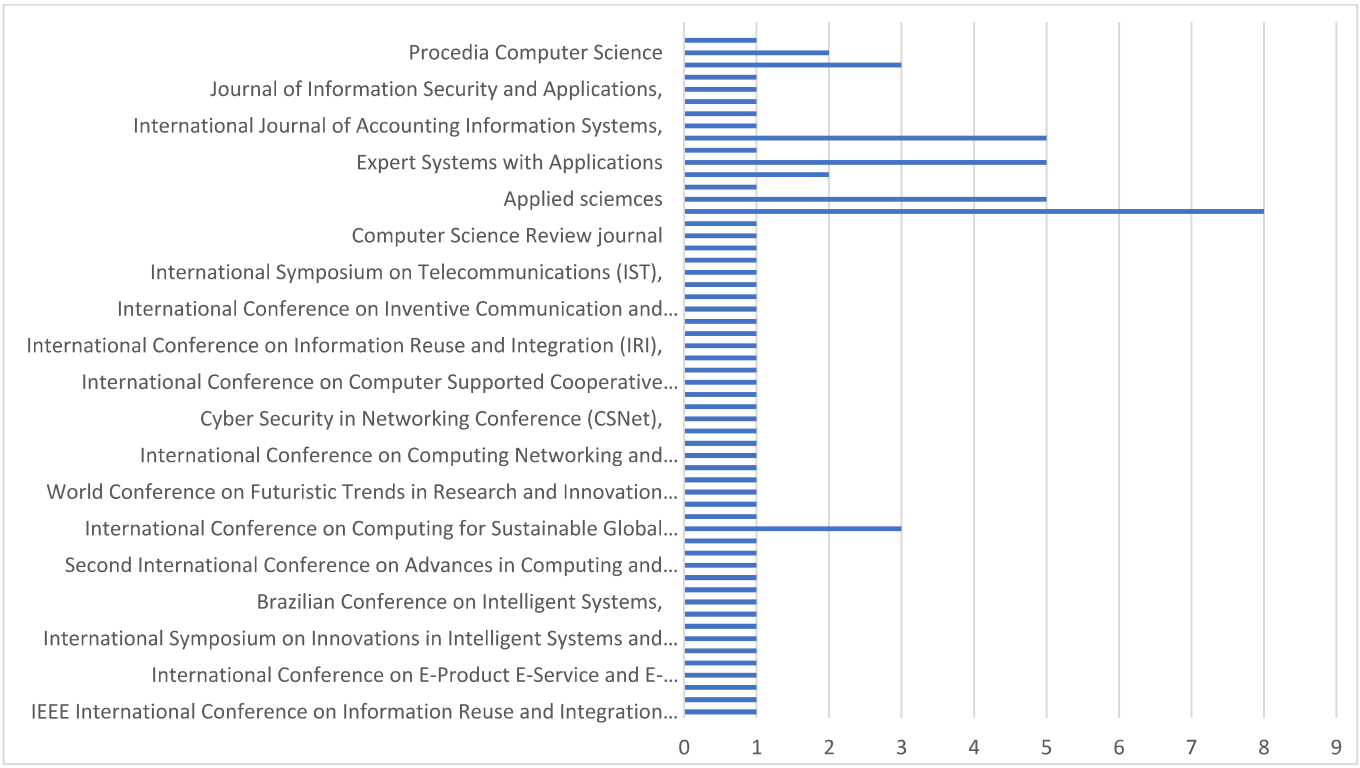

Figure 3. Number of articles per journal.

3.2. Synthesis Results

This section presents the results of the data synthesis to address the research questions based on the selected papers. Thus, in this section, the research questions designed for the SLR will be answered.

3.2.1. RQ1: What Are the Different Categories of Fraudulent Activities That Are Addressed Using ML Techniques

Fraudulent activities vary depending on industry sectors [1,4,27]. This section attempts to answer RQ1 by presenting different fraudulent activities that were addressed using ML techniques based on the selected articles. Based on the reviewed articles, fraudulent activities in the financial sector can be broadly categorized into credit card, mortgage, financial statement, and health care fraud. This can be further explained in the following subsections.

Credit Card Fraud

Credits are typically used to refer to electronic financial transactions made without the use of physical cash [28]. A credit card that is extensively used for online transactions is a small piece made up of thin plastic material with credit services and customer details [28–30]. Fraudsters use credit cards to make unlawful transactions that result in massive losses to banks and card holders [31]. Moreover, the invention of counterfeit cards has aided fraudsters in performing illicit transactions more easily. In general, it is regarded as illegitimate to use the card without the proper owners' authorization. By obtaining access to a certain account illegitimately, any transaction that is carried out is considered as fraudulent [29,30]. Credit card fraudulent activities can be divided into two aspects, namely, offline and online fraud [29]. In offline fraudulent activity, the fraudsters conduct their illicit transactions with stolen credit cards such as genuine card holders, while online fraudsters conduct their activities in online transactions through Internet Online fraud [29,30].

Financial Statement Fraud

Fraud in financial statements involves forging financial reports to claim that a company is more profitable than usual [3], avoid the payment of taxes, increasing stock prices, or obtaining a bank loan [32]. It can also be regarded as the confidential records generated by organizations that contain their financial records that comprise their expenses, profits made, income loans, etc. [33,34]. These statements also comprise some write-ups made by management for discussing business performances and predicted future tendencies [35–37]. Different financial records provide the financial reality of the organization, which indicates how successful the organization is and assists in checking if the organization is bankable [33,34]. In addition, financial statement fraudsters deceive the users of financial statements by correcting misstatements to make the organizations appear beneficial. The main purpose of the financial fraudulent statements is to enhance share prices, minimize tax liabilities, attract more investors as much as possible, and access personal bank loans among others [15].

Insurance Fraud

Insurance fraud can be defined as the act of misusing an insurance policy for gaining illegitimate benefits from an insurance business [38]. Usually, insurance is made to protect the organization's transactions or individual's transactions against any financial risks [33,34]. The main sectors of target by fraudulent insurance claims include healthcare [5,39,40] and automobile insurance companies [41,42]; although home and crop insurance fraudulent also occur [1], however, there is a paucity of the literature on both [16,43]. It has been estimated recently that the total cost of insurance fraud in the United States is over a billion USD yearly and it is finally passed on to consumers in the form of higher insurance premiums [11].

In order to cover the relevant costs of theft or accidental damages to a car, an agreement between the insurance provider and the insured person or organization is typically involved in automobile insurance claims [42,44]. Individual fraudsters are capable of committing fraudulent claims, and one method of committing fraud is through deception during the claims process [44]. Evidence of organized groups working together to conduct insurance fraud also exists [24]. Typically, these groups stage or fake incidents; in other cases, an accident may not have even occurred. Instead, the vehicles were brought to the scene [44]. Nevertheless, the majority of fraud cases are opportunistic frauds in a way that they are not planned; rather, an individual seizes the opportunity presented by such an accident by exaggerating the claimed statements or damages. Another popular insurance fraud is in the health sector [5,40]. Healthcare has grown to be a serious issue in contemporary society that is entangled with social, political, and economic concerns [39]. There is a significant financial expense associated with meeting the public demand for high-quality medical services and the technology required to provide them. Additionally, many low-income people and families rely on government-sponsored healthcare insurance programs for

support in order to pay for the steadily rising costs with respect to prescription medications and medical services [40].

Financial Cyber-Fraud

The term financial cyber fraud is a new term capturing the umbrella of crime committed over cyberspace for the sole purpose of illegal economic gain [45,46]. Financial cybercrime perpetrators are difficult to identify [47,48]. They purposely mask their activities to blend their actions with the normal behavior of any other customer or user of a website or financial service; however, when grouped together, the activity is more obvious in terms of its abnormality. As technical skills and advancements in technology are increasingly available to criminals, their tactics for committing criminal offenses become more difficult to combat. This symbiosis of financial crime and cybersecurity is leading financial institutions to use their in-house developed methods to protect their assets using tools such as real-time analytics and interdiction to prevent financial loss [49]. However, as models are showing signs of an inability to prevent and address these attacks [50], new methods must be developed and deployed across organizations to prevent further loss to their business, customer data, and their own reputation. The new methods deployed in the research community and industry include machine learning and deep learning models [47–50].

Other Financial Fraudulent Types

Apart from the above types of fraudulent activities committed in the financial sectors, other frauds are met in the financial domain, which includes commodities and securities fraud [32], mortgage fraud [5], corporate fraud, and money laundering [5]. Securities and commodities fraud is a dishonest practice that occurs when a person invests in a company based on given fake information [5]. A mortgage is a material misstatement made by a debtor at any stage of the application procedure when an underwriter relies on those facts to obtain a loan or credit [5]. It intentionally targets documents associated with a mortgage by modifying information during the mortgage loan application processes [7]. Another popular fraud is corporate fraud, which involves the falsification of financial documents by insiders to cover up any fraud or criminal activity [32]. Money laundering is another type of financial fraud in which fraudsters try to change the source of illegal money by convincing criminals to turn their dirty money into legitimate money [1,5]. Money laundering has a major influence on society because it is the primary method in which other crimes, such as funding terrorism and trade-in weapons, are accomplished [4,5]. Another popular financial crime is cryptocurrency fraud [51]. This type of fraud systematically provides fake investments to naïve users in order to defraud them [35,52]. The main idea of this is to entice innocent individuals with the promise of significant gains from their investments [34,53]. Table 5 show the different types of financial fraud.

Table 5. Types of financial fraud.

Fraud Type	Description	Technique Used	References	No. of Reference
Financial Statement Fraud	This is a corporate fraud such that the financial statements are illegitimately modified to allow the organizations to look more beneficial.	Support Vector Machine	[33,54–56]	20
		Clustering based method	[37,56]	
		Decision Tree	[33,57–60]	
		Logistic Regression	[35]	
		Naïve Bayes	[33,61]	
		Artificial Neural Network	[33,40,62,62–64]	

Table 5. *Cont.*

Fraud Type	Description	Technique Used	References	No. of Reference
Credit Card Fraud	Illegitimate use of the card without proper owners' authorization	Support Vector Machine	[18,57,65–67]	32
		Fuzzy logic	[68,69]	
		Clustering based method,	[70,71]	
		Artificial Neural Network	[70,72]	
		Hidden Markov model	[20,32,73]	
		Decision Tree	[16,21,74]	
		Genetic Algorithm	[19,75]	
		Artificial Neural Network	[30,61,76,77]	
		Naïve Bayes	[5,28,69,78]	
		Logistic Regression	[78]	
		Random Forest	[29,69,79],	
Health Insurance Fraud	Fraudulent claims by individuals or organizations to support the relevant expenses of theft or accidental damages.	Support Vector Machine	[80]	5
		Artificial Neural Network	[81]	
		K Nearest Neighbors	[82]	
		Naïve Bayes	[83],	
		Clustering-based method,	[40]	
Auto Insurance Fraud	Fraudulent claims by an individual to get health insurance profits.	Support Vector Machine	[41,42],	3
		K Nearest Neighbors	[84]	
Cyber Financial fraud	Financial fraudulent activities through cyber space	Artificial Neural Network	[47–50]	3
		SVM	[50,85]	2
Others	Other frauds that are faced in the financial domains include commodities and securities fraud [32], mortgage fraud, corporate fraud, and money laundering.	Support Vector Machine	[42]	5
		Decision Tree	[42]	
		Fuzzy logic	[69]	
		Clustering-based method,	[51]	
		Hidden Markov model	[86]	

3.2.2. RQ2: What Are the ML-Based Techniques for Financial Fraud Detection Employed in the Literature?

Machine learning (ML) is referred to as analytic techniques that find specified patterns without the manual guidance of an expert [87,88]. Financial fraud detection has been widely studied using ML methods by many researchers [17,89,90]. This includes SVM, ANN, HMM, KNN, Decision Tree, etc. Therefore, to answer the above research question (RQ2), this section presents different popular ML methods that were used for financial fraud detection based on the selected articles in the review. A detail explanation of the ML techniques used for detecting financial fraudulent activities is provided in the following sub-section.

Support Vector Machine (SVM)

SVM is a supervised ML method that seeks a maximum margin hyperplane for classifying input training data into two categories [41,66]. SVM is capable of classifying new data points based on a labeled training set for each class [68]. Based on the reviewed literature, several researchers investigated SVM techniques for fraud detection [65,66,80]. For example, Rajak and Mathai [65] introduced a hybrid technique based on SVM and the fusion Danger theory for the fraudulent detection approach. Based on the experimental results, the authors showed that this study outperformed the existing approaches in terms of time complexity and F-measure. Francis et al. [80] used the SVM technique to propose fraud detection by investigating an automated medical bill architecture. This research method aims provide a quick response for detecting medical fraud in real time. Experimental results showed that the model performed better compared to the previous approaches. Xu and Liu [66] applied optimized SVM to detect fraudulent activities in an online credit card.

In this approach, the authors examined the model performance using commercial banks' business datasets. The results demonstrated that the SVM outperformed other compared models, thereby confirming its viability.

Mareeswari and Gunasekaran [57] presented an approach for detecting fraudulent behaviors in credit cards based on SVM techniques. The authors combine a hybrid SVM and spike detections to address the limitations of existing methods. It is proven that the result of the proposed approach outperformed existing methods. Gyamfi and Abdulai [67] developed an SVM-based supervised learning approach to distinguish between illegal and legal customer behaviors in credit card transactions. To increase the accuracy detection rate, the authors combined SVM, logic, and linear regression methods. A study by Sundarkumar et al. [41], which utilized a one-class SVM-based under-sampling technique, was presented to enhance fraud detection in insurance industries. A study in [66] presented a fraudulent detection method for credit card transactions by integrating OSVM with a deep learning approach. The method was assessed using real-world datasets, and the results showed promising results.

Fuzzy-Logic-Based Method

Fuzzy logic (FL) is an effective conceptual framework for addressing the issue of representing the data in a context of uncertainty and ambiguity [69,91]. It is a logic that shows that methods of thinking are not accurate but estimated [70]. The Fuzzy combinations offer effective concepts for handling complex modeling in a new and better way [92]. Several methods based on the FL have been used for fraud detection. To detect anomalous behaviors in credit card transactions, the FUZ-ZGY hybrid model, based on the fuzzy and Fogg behavioral models, was introduced in [69]. A system based on fuzzy logic was employed to track the historical activities of the merchant, and the Fogg behavioral method was employed to characterize the customer's behavior along two different but related dimensions: the ability to commit fraud and motivation. Another fuzzy-based method was proposed in [68] to detect fraud in credit cards by categorizing the fraud transactions and non-fraud transactions with decreased false positives. The method used fuzzy c-means clustering and the ANN model. The model was evaluated on synthetic data and the results showed that the combination of clustering techniques and learning mechanisms helps in reducing false positives.

A study in [69] provides a fuzzy logic-based fraud-detection method in the banking system. This work was able to improve detection accuracies in classifying fraudulent and non-fraudulent activities in banking transactions. The authors defined some rules for fuzzy logic based on the experience of the experts to improve detection accuracies. This approach was improved in [52] by constructing fuzzy rules using fuzzy logic for the improved detection of fraud transactions. Pradeep et al. [91] introduced rule-based technique using a firefly algorithm and the threshold-accepting method to differentiate between fraudulent and non-fraudulent transactions based on their financial activities. Hajek [92] designed a fuzzy-rule-based approach for detecting financial fraud by integrating a rule-based approach with genetic feature selections. The approach was capable in achieving good performances by using the feature selection method to remove irrelevant attributes and performing a fuzzy unordered rule induction.

Hidden Markov Model (HMM)

The HMM is a dual embedded random method often used to perform more complex random processes better than the traditional Markov model [19]. Based on the reviewed literature, several methods have employed the HMM technique for financial fraud detection. Khan et al. [93] used an HMM-based method for detecting the card owners' behaviors through the observed incoming transaction. The author used a clustering approach to differentiate between fraudulent and non-fraudulent patterns using data conglomeration of regions of the parameter. Agrawal et al. [19] introduced a hybrid method by integrating HMM and Genetic algorithms (GA) for identifying credit card fraudulent transactions. The

approach employed the HMM to preserve previous transaction logs and GA to compute the threshold value for clustering incoming transactions into several clusters. The authors showed that the method used is more effective for credit card fraud detection. A similar approach was proposed in [86] to achieve internet banking fraud detection by disclosing the right users and monitoring their illicit behaviors.

A method in [73] utilized HMM to address the limitation of existing fraud-detection methods when the transaction is accomplished in credit card operations. The findings of the study suggested that HMM is capable of improving fraud detection and as well minimizing false-positive rates. Wang et al. [94] conducted a simulation experiment using HMM and K-means methods for improving bank fraud detection. The result demonstrated the ability of the model to effectively discover fraud in bank transactions. A similar approach proposed in [20] used an HMM-based technique, which improves the efficiency and accuracy of credit card fraud detection. To determine the clusters' closest centroids and integrate them into a single group, the authors also employed the clustering technique based on the K-means method.

Artificial Neural Network (ANN)

ANN is an information-processing technique inspired by biological neural network behavior [76]. ANN is very powerful when there is the availability of a large volume of data [95–97]. Several ANN-based methods have been proposed for fraudulent detection in the financial sector. Using an ANN-based method, Srivastava et al. [30] investigated credit card fraud detection on the trader's side. The proposed method connects the merchant with payment gateways. The payment gateway serves as a medium between the merchant, who owns the customer's credit card information, and the fraud-detection model. To identify credit card fraud, Ghobadi and Rohani [77] developed a hybrid model based on a Cost-Sensitive Neural Network. The findings demonstrated that the suggested model increased the detection rate and reduced false negative costs. Randhawa et al. [28] proposed research for discovering fraud in credit cards transaction based on ML methods. The study used several ML algorithms, including ANN models.

A study based on NN was introduced in [76] for the fraudulent transactions detection in credit cards to improve the security and accuracy of automatic credit card transactions. Ravisankar et al. [33] introduced financial fraud detection using a multilayer feed forward neural network (MLFF). An ANN technique based on deep reinforcement learning (DRL) was used in [97] for detecting fraud in the banking sector. The authors discussed several interesting facts about DRL, which show that it performs competitively when compared to other approaches.

KNN Algorithm

The K-nearest neighbors (KNN) algorithm is a convenient, straightforward supervised ML technique that is powerful in addressing both regression and classification processes [62,98]. The class label is usually determined by the KNN model using a small set of the nearest samples. The KNN model is a type of non-parametric model that is used for both classification and regression tasks and that can locate similar neighborhoods that are closest to a given sample point in a dataset and create a new sample point based on the distance between two samples of data [70,99]. Although it worked well on many datasets, the performance of this technique is likely compromised by unbalanced datasets [78]. The Euclidean distance [100] is one of the most well-known techniques for calculating distance.

However, some recent approaches have been introduced for financial fraud detection. For example, Malini and Pushpa [70] proposed a credit card detection approach by using two different methods: the KNN model and the outlier detection model. The experimental results showed that the KNN model is more effective for fraudulent detection in credit cards. Awoyemi et al. [78] used the KNN algorithm to investigate credit card transactions for detecting fraudulent behaviors. The authors used a credit card dataset proposed by European cardholders. The finding demonstrated that the K-Nearest Neighbor performed

better than other existing techniques. Badriyah et al. [84] introduced a KNN-based approach for fraud detection in auto insurance. The approach comprises three methods, namely, distance-based, density-based, and interquartile range in car insurance data. The proposed work considers the impact of feature selection methods on accuracy scores. Similar methods have been introduced in [72] to discover the anomaly fraudulent transaction by integrating the KNN technique and Chi-Square Automatic Interaction Detection (CHAID) to enhance the performance of fraudulent transactions.

Bayesian Method

The Bayesian model (BN) is a particular type of graphical model that takes into account both independent and conditional relationships between various variables [101]. A directed graph's nodes and edges are used by the BN. The Bayesian model is a particular type of graphical model that takes into account both independent and conditional relationships between various variables [96]. A directed graph's nodes and edges are used by the BN. This model is very powerful in searching anonymous probability computations [56]. Based on the reviewed literature, we explored different papers on the two main types of Bayesian methods, namely, the Bayesian belief network and Naive Bayes (NB). NB is an ML model that is based on the Bayes theorem and is used to predict membership probabilities per class [101]. It predicts a given data point label based on the probability that belongs to a particular category [56]. Some researchers utilized the NB model for financial fraud detection. For example, Deng [101] utilized the NB model to produce a fraud detection approach in a financial transaction. The experiment was performed in a dataset that contains both normal and abnormal financial statements. The results showed the effectiveness of the proposed model in fraud detection. Richter and Herland [81] utilized the NB algorithm to address fraudulent transactions in the health sector based on medical procedure records. The research is aimed to classify the supplier's behavior with respect to whether it is anomalous or not. To better perform fraud detection, Hajek and Henriques [33] proposed a method for intelligently detecting fraudulent financial documents by gathering specific features from financial reports. In the experiment, the authors took into account several linguistic and financial management discussion variables related to financial information. The result showed that the proposed model outperformed other existing ML methods. Some studies also used Bayesian networks to predict fraudulent financial transactions [33,63].

Decision Tree

A decision tree (DT) is an ML technique that is used for creating decision support tools in the trees of inner nodes, which represent binary options over the features [69]. For many years, there have been several methods based on the decision trees that are used to detect financial frauds. Devi and Kavitha [78] developed a DT base method to classify credit card transactions as normal or suspicious data. The method was evaluated with different accuracy metrics. The results indicated that DT performed better than the existing approaches with a high degree of accuracy. A study was conducted in [79] for auto fraud detection by using an ML technique. The authors compared three different methods including NB, DT, and RF methods, and the result proved that DT outperformed other methods.

Kho and Vea [67] investigated the transaction behavior of credit cardholders to differentiate between normal and abnormal transactions. The authors applied different ML algorithms such as Random Tree (RT) and NB. The methods were evaluated on two synthetic datasets. The results of the evaluation show that RT performed better compared to other methods. Another similar approach to detect fraud in the auto insurance sector was introduced in [42] by utilizing an adaptive oversampling method. The method was able to delete the imbalance classes from the insurance datasets.

Genetic Algorithm

The genetic algorithm (GA) draws inspiration from natural evolution [35]. It uses a set of suggestions that are often represented as binary strings called chromosomes to search for the best solutions [75,100]. Genetic programming belongs to the area of evolutionary algorithms that widens the use of genetic methods to support the exploration of computer programs' space [76]. Many approaches used GA in the literature. For example, in [35] Gupta and Gill employed GA for financial fraud detection in companies. Benchaji et al. [71] present a new technique for fraud detection in credit card transactions to address the issue of the conventional methods in detecting minority class objects in the imbalanced datasets using the K-means and GA techniques. The authors first employed the K-means method to group and classify the minority instances; then, they applied the GA method in each group to create new instances in order to obtain a new training dataset. In addition, the GA algorithm was also utilized by Özçelik et al. [61] to provide a solution to these problems of detecting fraudulent credit card transactions. The study was applied to a real-life application project using transactional data from the real world.

Ensemble Methods

The ensemble method is a meta-algorithm that combined manifold intelligent techniques into one predictive technique [33,98]. The main goal of ensemble methods is to address the weakness of the individual models relative to stronger models [98]. Every ensemble technique follows a different purpose; for example, boosting manages to decrease bias, bagging tries to decrease variance and stacking attempts to enhance predictions [33,98]. Several ensemble techniques were used in several studies [33,98]. Based on the reviewed articles, random forest (RF) is the most commonly used ensemble technique in the literature [33,98]. In particular, it outputs the median prediction for regression tasks and the mode of the classes of the single trees in classification problems. A recent study showed that RF performed better than other compared methods [64].

Bagging, which is also called Bootstrap Aggregating (BA), generates several samples from the training instances with replacements. Based on the reviewed papers, several research studies applied bagging techniques for financial fraud detection [33]. By changing the distribution of the training dataset based on the accuracy of the predecessor, boosting aims to train weak learners serially [28]. One of the popular boosting techniques is AdaBoost, [28]. AdaboostMI is a multi-instance Adaboost that was used in [33] to repetitively execute different distributions of SVM throughout the training dataset and then combined the classifiers into a separate hybrid classifier. Stacking is another ensemble ML method that aggregates different classification or regression models [2]. Unlike bagging, which uses samples of the training data, stacking uses the entire dataset and uses models that are typically different from one another.

Clustering Based Methods

Clustering is an unsupervised learning method that involves grouping identical instances into the same sets [70,102]. Although Clustering techniques are popular in financial fraud detection, they were, however, implemented considerably less than classification techniques in the reviewed articles [5,102,103]. Glancy and Yadav [56] used text-mining hierarchical clustering to design a financial transaction fraud-detection model. Glancy and Yadav [56] proposed an approach for financial fraud detection using the text dimension reduction method and document clustering. The authors used the SVDs technique to achieve text dimension reduction. The dual GHSOM technique is created to detect the non-fraud-central spatial hypothesis [18]. The model is capable of detecting the topological patterns of fraudulent financial transactions. Deng and Mei [102] integrated K-means clustering and SOM to design a clustering-based fraud-detection method. To address the uncertain clustering borders of nodes, which is one of the shortcomings of the SOM, this model additionally applied K-means clustering.

Logistic Regression

Logistic regression (LR) techniques are mainly applied in binary and multi-class classification problems [35,78]. It operates by performing regression on a set of variables. It is typically a useful technique for describing patterns and clarifying connections between numerous dependent binary variables. In line with a review article by Abbasi et al. [104], the logistic regression method is one of the most used machine learning (ML) techniques for detecting financial misstatement models. Based on that review work, the majority of the studies used LR techniques for financial fraud detection. An appropriate technique for identifying characteristics related to fraudulent transaction detection using LR was proposed by Peng and You. [81] after reviewing published data. The authors compared the predictive ability of the suggested method against other detecting methods. Table 6 presents ML techniques used for financial fraud detection.

Table 6. ML techniques used for financial fraud detection.

Techniques	Short Description	No. of Articles	References
SVM	A classification method used in linear classification	10	[18,41,42,54,57,65,66,66,80,80]
HMM	A dual embedded random process used to provide more complex random processes	8	[19,20,73,74,86,89,94,103]
ANN	Amulti-layer network that works similar to human thought	10	[28,30,33,61,62,75–77,81,97]
Fuzzy Logic	A logic that indicates that methods of thinking are estimated and not accurate.	5	[68,69,91,92]
KNN	It classifies data according to their similar and closest classes.	7	[60,70,72,78,84,98,99]
Decision Tree	A regression tree and classification method that is used for decision support	5	[29,44,54,57,67]
Genetic Algorithm	It searches for the best way to solve problems concerning the suggested solutions	3	[35,61,71]
Ensemble	Meta algorithms that combined manifold intelligent technique into one predictive technique	8	[2,29,33,64,69,79,98,105],
Logistic Regression	They are mainly applied in binary and multi-class classification problems.	8	[35,78,81,104,106–108]
Clustering	Unsupervised learning method which involve grouping identical instances into the same sets	6	[5,18,56,102,103]
Random Forest	Classification methods that operate by combining a multitude of decision trees	7	[21,29,67,79,109–111]
Naïve Bayes	A classification algorithm that can predict group membership	11	[5,28,31,33,61,67,78,83,101,101,112]

3.2.3. RQ3: What Are the Performance Evaluation Metrics Used for Financial Fraud Detection Using Machine Learning Methods

To evaluate the performance of a model, the evaluation metric is very important in financial fraud detection [38,40,84]. However, there are no specific evaluation measures that are strictly used for evaluating ML techniques for fraud detection [38,72]. In recent times, several performance evaluation metrics have been employed by different researchers that include accuracy, precision, recall, F1 measure, false-negative rate (FNR), the area under the curve (AUC), specificity, etc. In this section, we present evaluation metrics that have been employed in the reviewed papers based on the selected articles. Table 7 shows the formulas of the different performance measures.

Table 7. Performance evaluation metrics.

Metrics	Formula	References
Accuracy	$Accuracy = \frac{(TN+TP)}{(TN+FN+FP+TP)}$	[18,21,29,33,38,40–42,62,66,69,72,78,82–84,110,113–117]
Precision	$Precision = \frac{TP}{(TP+FP)}$	[21,39,40,67,69,77,78,94,113,118]
Recall/Sensibility/TPR)	$Recall = \frac{TP}{(TP+FN)}$	[21,39,40,67,69,77,94,113,118]
F-measure(F1)	$F1 = 2 * \frac{Recall*Precision}{(Recall+Precision)}$	[21,31,33,65,84,113]
Specificity (TNR)	$Specificity = 1 - FP$	[42,70,77,78,82,94,113]
AUC	AUC = the area under ROC curve	[35,113]
Others		[65–67,74,119]

The model's accuracy measures how many of the model's overall predictions were accurate, whereas precision measures how accurate the model's positive predictions were [42,69,82]. The percentage of positive cases that the classifier properly detects is known as the recall of a classifier, also known as sensitivity [21,67,113]. The classifier's specificity is measured as a ratio of correctly classified negative samples to all negative samples [78,82,113]. While it makes sense to increase both recall and precision, the two metrics have an inverse relationship. Lower recall may arise by forcing a higher precision, and vice versa [78,82,113]. This is referred to as the recall/precision trade-off. The F-score, which is the harmonic mean of the precision and recall is a better metric to maximize instead [21,56,113]. It is possible to track the performance change in terms of the trade-off between certain measures such as recall and precision by tweaking the decision threshold value for a classifier. This process of evaluating all confusion matrices that can result from changing the decision threshold is known as parametric evaluation.

Another metric is the Receiver Operating Characteristic (ROC) curve and the precision–recall curve [35,113]. The area under the ROC curve (AUC) can be used to compare classifiers, with perfect classifiers possessing an AUC equal to 1 and purely random classifiers having an AUC equal to 0.5. Different distance measures, such as the mean squared error (MSE), Euclidean distance, Manhattan distance, and others, are used by clustering algorithms to quantify the similarity or dissimilarity of samples or observations. These algorithms group similar data points based on relevant features. Table 7 presents the formulas of performance measures with their corresponding references and the number of studies for each metric. As for the FPR and FNR, the lower the value of the performance measures, the more improved the generalization ability of models. In this table, TP, TN, FP, and FN stand for the true positive, true negative, false positive, and false negative, respectively.

3.2.4. RQ4 What Are the Gaps and Future Research Direction in Machine-Learning-Based Fraud Detection

This section attempts to answer RQ4, which aims to present research gaps and future directions of the field. Thus, to determine the limitations and explore the gaps and future work opportunities in this field, we synthesized the reviewed articles as discussed in the following subsections.

Imbalanced Dataset

Virtually, most financial transaction datasets comprise millions of transactions, and all of them share a common issue, namely imbalanced datasets [1,5]. On other hand, the number of fraudulent financial transactions is far fewer than non-fraudulent ones. [33]. This issue is generally caused as a result of the fact that the rate of actual fraud transactions out of all transactions is nominal [2]. The problem of imbalanced data distribution generally affects the efficiency of machine learning models [87]. Therefore, training models for detecting fraudulent activity that is very minimal requires extra consideration.

To address the issue of imbalanced data, some studies applied oversampling approaches [12,120]. The others attempt to introduce approaches that may effectively work with extremely imbalanced data. For example, Li et al. [86] and Perols [61] utilized imbalanced and left balancing datasets based on the oversampling method for future work. Additionally, Lin et al. [94] avoided oversampling as it may lead to choice-based sample biases. Essentially, there were no studies that are applied under-sampling methods among the reviewed articles. Moreover, the only applied oversampling method was the SMOTE (Synthetic Minority Oversampling Technique) [121]. Because of this, it could be deduced that future studies could consider employing other oversampling techniques, as well as under-sampling techniques.

Data Size Feature Vectors

Some research works established that the dataset's size is a limitation of their studies. For instance, Jan [94] showed that the financial markets in Taiwan are smaller than Europe, Japan, and China in terms of its scope and size. Moreover, in Taiwan, the number of registered companies is relatively small in scale. Accordingly, the data's size is one of the major problems in other nations. Therefore, if the issue of the datasets can be resolved, an improved and more efficient ML approach that could identify fraudulent financial activity can be achieved.

On other hand, the majority of the studies in the reviewed literature noted that the performances of the detection model can be enhanced by improving the input vectors. Future works could combine data from other sources, such as financial social media sites such as Seeking Alpha, numerical information from financial documents, and the transcripts of earnings calls, to produce more relevant feature vectors. In addition, if the textual data could be considered in the model's design process, then exploring emerging methods of converting textual contents into vectors, such as Word2Vec, and BERT, such as the Doc2Vec algorithm, could benefit further research.

Unstructured Data

Recently, several studies investigated different kinds of unstructured data [122] such as vocal inputs and textual data. However, to achieve remarkable results, unstructured data exploration in the financial fraud detection domain requires more attention. We expect future research to look into the text sources from financial statements. In addition, future research could also explore the use of new data mining techniques, particularly word-embedding methods such as Doc2Vec, Word2Vec, and BERT, to convert financial texts into feature vectors that can be used to create machine learning models.

Machine-Learning-Based Technique

Classifying the machine learning techniques used for financial fraud detection is an effective method for determining the appropriate methods for this research domain [123]. Research gaps can be identified by investigating why particular methods were selected and why others were not given more attention. Based on the literature, many learning algorithms that have received more attention and are used in other fields have not been popularly applied in financial fraud detection [5]. For instance, active learning can address the issue of insufficient data and improve the learning cost, incremental learning can dynamically add sample data to accuracy, and transfer learning can use the knowledge acquired from learning tasks to enhance the learning effect on other tasks. In future, these learning methods can be given more attention in financial fraud detection.

Another future direction that is worth considering in machine learning is the concept of drift [124]. Many times, due to the continuous development and evolution of financial fraud, the accuracy of the algorithms diminishes. This phenomenon is called concept drift, [124], which is a problem that scholars have been attempting to address in the area of ML. Although the concept drift issue can be addressed by applying new datasets for

training periodically to update the algorithms, it is clear that the overhead is large, and the model performance between two training cycles cannot be certain.

4. Discussion

In this section, the content of the SLR is highlighted, which includes popular financial fraud detection and machine learning techniques used in detection methods. We categorized the findings of ML techniques and the fraud type in this SLR based on their frequency of usage. From the review, it can be observed that out of all ML approaches discovered in this study, the most popular ones from 2010 to 2021 are summarized. As a result, we identified that the NB algorithm is the most popular technique used for identifying fraudulent activities in the financial sector followed by the SVM and ANN with 11, 10, and 10 articles, respectively. This shows that the NB, SVM, and ANN are the most popular machine learning techniques used for the financial fraud detection based on the reviewed literature. Figures 4 and 5 shows the frequency distribution of the machine learning techniques used for fraud detection and the financial fraud types addressed in the reviewed articles.

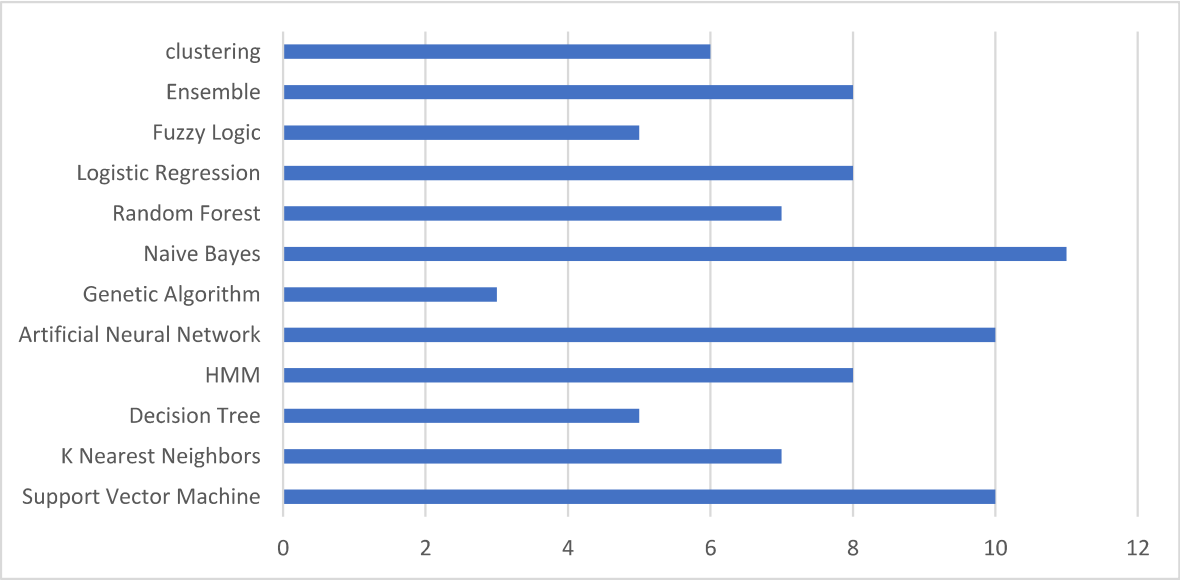

Figure 4. Frequency of the machine learning methods used for fraud detection.

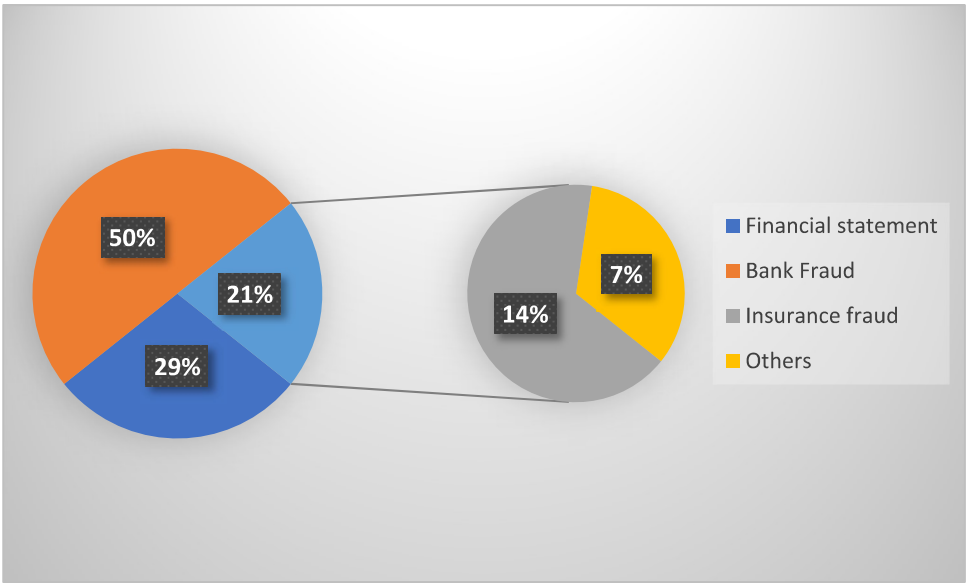

Figure 5. The frequency of the different fraud types.

Based on the findings of our study, we categorized financial fraud into four different categories, such as financial statements bank fraud, insurance fraud, and others. The number of reviewed papers and the type of fraud are provided in Figure 4. It can be shown that out of the review papers, 50% articles focus on bank fraud, 29% addressed financial statements, and 21% focused on insurance fraud and other fraud at 14% and 7%, respectively. This show that the majority of studies have a significant focus on bank fraud and financial statement fraud, while insurance frauds that include health insurance and auto insurance are not as frequently identified in the reviewed articles. It also shows that the two most common frauds include credit card and financial fraud, which appear to be the most prevalent types of fraud. On the other hand, our review did not cover money laundering, stock, commodities, or mortgage fraud for some reason. One of the factors is the difficulty in acquiring these data and the inability to reveal results if they are related to sensitive subjects.

5. Limitation and Threat to Validity

In this SLR, various ML techniques and fraud types were identified. We develop our protocols to promote external and internal validity as much as possible while answering the RQs. However, there are still some limitations and validity threats that can be encountered and mentioned here.

1. This SLR is only limited to conference and journal papers that discuss machine learning (ML) in the context of detecting financial fraud. By using our search approach in the early stages of the review, several non-relevant research papers were identified and excluded from this review. This ensures that the selected research papers satisfied the criteria for the study. However, it is believed that using more sources, such as additional source books, would have further enhanced this review.
2. Although major databases were taken into consideration when exploring the research articles, there may be other digital libraries with relevant studies that were overlooked. We compared search terms and keywords to a well-known list of research studies to mitigate this limitation. However, some synonyms may be overlooked when searching for the keywords. The SLR protocol has been revised to address this problem by ensuring no essential terms are left out.
3. We restricted our search to only English-language articles. Thus, this results in linguistic bias because some related papers in this field of study may exist in other languages. However, fortunately, all the gathered papers in this study were written in English. As such, we have no language bias.

6. Conclusions

Financial Fraud can be committed in different financial aspects such as insurance, banking, taxation, and corporate sectors [3]. Recently, financial fraud has become increasingly worrisome among companies and industries [4]. Despite several efforts to eradicate financial fraud, its persistence adversely affects the economy and society as very large amounts of money are lost to fraud every day [6]. With the advent of artificial intelligence, machine-learning-based approaches can be used intelligently to detect fraudulent transactions by analyzing a large number of financial data. In this paper, we presented a study that systematically reviewed and synthesized the existing literature on ML-based fraud detection. In particular, this paper adopted the Kitchenham methodology, which uses well-defined protocols to extract, synthesize, and report results. Several studies have been gathered based on the specified search strategies for popular electronic libraries. After the inclusion/exclusion criteria, 87 were selected. In this review, popularly used ML techniques for fraud detection, the most common fraud type, and the evaluation metrics are summarized. Based on the reviewed articles, results showed that SVM and NN are the popular ML algorithms used for fraud, and credit card fraud is the most popular fraud type in the literature. The paper finally presented the key issues, gaps, and limitations in the area of financial fraud detection and suggests areas for future research. We identified

gaps in the research by examining unexplored or less studied algorithms. Previous studies in financial fraud detection focused on supervised classification and regression methods, such as SVM, neural networks, and logistic regression. The use of ensemble methods that take advantage of multiple algorithms to classify samples is a rising trend in the field. Interestingly, we discovered that unsupervised learning approaches, such as clustering, were less employed in the present literature. Clustering is beneficial for investigating latent relations and resemblances. In addition, since there are a small number of fraud cases that have to be identified, clustering could be effective. We recommend that future studies pay more attention to unsupervised practices, such as anomaly detection, which can uncover new insights. Additionally, another avenue for future research would be to use emerging text-mining techniques and word-embedding techniques such as Word2Vec, Doc2Vec, or BERT to transform financial texts into vectors of features, which will then be used to build machine learning models.

Author Contributions: A.A., conceptualization, data curation, formal analysis, resources, visualization, writing—original draft, and writing—review and editing; S.A.R., supervision, resources, conceptualization, project administration, and funding acquisition; S.H.O., supervision, data curation, formal analysis, resources, and visualization; T.A.E.E., conceptualization, data curation, formal analysis, and funding acquisition; A.A.-D., resources, data curation, visualization, software, and validation; M.N., writing—original draft, data curation, validation, and funding acquisition; T.E., resources, data curation, and visualization; H.E., writing—review and editing; A.S., investigation and formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deanship of Scientific Research at King Khalid University through Large Groups (Project under grant number (RGP.2/49/43)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups. (Project under grant number (RGP.2/49/43)).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hilal, W.; Gadsden, S.A.; Yawney, J. Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Syst. Appl.* **2021**, *193*, 116429. [CrossRef]
2. Ashtiani, M.N.; Raahemi, B. Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review. *IEEE Access* **2021**, *10*, 72504–72525. [CrossRef]
3. Albashrawi, M. Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015. *J. Data Sci.* **2016**, *14*, 553–570. [CrossRef]
4. Choi, D.; Lee, K. An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation. *Secur. Commun. Netw.* **2018**, *2018*, 1–15. [CrossRef]
5. Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decis. Support Syst.* **2011**, *50*, 559–569. [CrossRef]
6. Ryman-Tubb, N.F.; Krause, P.; Garn, W. How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Eng. Appl. Artif. Intell.* **2018**, *76*, 130–157. [CrossRef]
7. Al-Hashedi, K.G.; Magalingam, P. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Comput. Sci. Rev.* **2021**, *40*, 100402. [CrossRef]
8. Chaquet-ulldemolins, J.; Moral-rubio, S.; Muñoz-romero, S. On the Black-Box Challenge for Fraud Detection Using Machine Learning (II): Nonlinear Analysis through Interpretable Autoencoders. *Appl. Sci.* **2022**, *12*, 3856. [CrossRef]
9. Da'U, A.; Salim, N. Recommendation system based on deep learning methods: A systematic review and new directions. *Artif. Intell. Rev.* **2019**, *53*, 2709–2748. [CrossRef]
10. Zeng, Y.; Tang, J. RLC-GNN: An Improved Deep Architecture for Spatial-Based Graph Neural Network with Application to Fraud Detection. *Appl. Sci.* **2021**, *11*, 5656. [CrossRef]
11. Delamaire, L.; Hussein, A.; John, P. Credit card fraud and detection techniques: A review. *Banks Bank Syst.* **2009**, *4*, 57–68.

12. Zhang, D.; Zhou, L. Discovering Golden Nuggets: Data Mining in Financial Application. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2004**, *34*, 513–522. [CrossRef]
13. Raj, S.B.E.; Portia, A.A. Analysis on credit card fraud detection methods. In Proceedings of the 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), Tirunelveli, India, 18–19 March 2011; pp. 152–156. [CrossRef]
14. Phua, C.; Lee, V.; Smith, K.; Gayler, R. A Comprehensive Survey of Data Mining-based Fraud Detection Research. *arXiv* **2010**, arXiv:1009.6119.
15. West, J.; Bhattacharya, M. Intelligent financial fraud detection: A comprehensive review. *Comput. Secur.* **2016**, *57*, 47–66. [CrossRef]
16. Abdallah, A.; Maarof, M.A.; Zainal, A. Fraud detection system: A survey. *J. Netw. Comput. Appl.* **2016**, *68*, 90–113. [CrossRef]
17. Popat, R.R.; Chaudhary, J. A Survey on Credit Card Fraud Detection Using Machine Learning. In Proceedings of the 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–12 May 2018; pp. 1120–1125.
18. Gyamfi, N.K.; Abdulai, J. Bank Fraud Detection Using Support Vector Machine. In Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 1–3 November 2018; pp. 37–41.
19. Carneiro, E.M.; Dias, L.A.V.; Da Cunha, A.M.; Mialaret, L.F.S. Cluster Analysis and Artificial Neural Networks: A Case Study in Credit Card Fraud Detection. In Proceedings of the 2015 12th International Conference on Information Technology-New Generations, Mumbai, India, 11–14 December 2011; pp. 122–126. [CrossRef]
20. Iyer, D.; Mohanpurkar, A.; Janardhan, S.; Rathod, D.; Sardeshmukh, A. Credit card fraud detection using Hidden Markov Model. In Proceedings of the 2011 World Congress on Information and Communication Technologies, Mumbai, India, 11–14 December 2011; pp. 1062–1066.
21. Patil, S.; Nemade, V.; Soni, P. ScienceDirect Predictive Modelling For Credit Card Fraud Detection Using Data Analytics. *Procedia Comput. Sci.* **2018**, *132*, 385–395. [CrossRef]
22. Mohammadian, V.; Navimipour, N.J.; Hosseinzadeh, M.; Darwesh, A. Comprehensive and systematic study on the fault tolerance architectures in cloud computing. *J. Circuits Syst. Comput.* **2020**, *29*, 2050240. [CrossRef]
23. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Keele University: Keele, UK, 2007; p. 65.
24. Pourhabibi, T.; Ong, K.-L.; Kam, B.H.; Boo, Y.L. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst.* **2020**, *133*, 113303. [CrossRef]
25. Marcotte, P.; Petrillo, F. Multiple Fault-tolerance Mechanisms in Cloud Systems: A Systematic Review. In Proceedings of the 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Berlin, Germany, 28–31 October 2019; pp. 414–421.
26. Isong, B.E.; Bekele, E. A systematic review of fault tolerance in mobile agents. *Eng. Appl.* **2013**, *2*, 111–124. [CrossRef]
27. Nassif, A.B.; Abu Talib, M.; Nasir, Q.; Dakalbab, F.M. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access* **2021**, *9*, 78658–78700. [CrossRef]
28. Randhawa, K.; Loo, C.K.; Seera, M.; Lim, C.P.; Nandi, A.K. Credit Card Fraud Detection Using AdaBoost and Majority Voting. *IEEE Access* **2018**, *6*, 14277–14284. [CrossRef]
29. Bhattacharyya, S.; Jha, S.; Tharakunnel, K.; Westland, J.C. Data mining for credit card fraud: A comparative study. *Decis. Support Syst.* **2011**, *50*, 602–613. [CrossRef]
30. Srivastava, A.; Yadav, M.; Basu, S.; Salunkhe, S.; Shabad, M. Credit card fraud detection at merchant side using neural networks. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 667–670.
31. de Sá, A.G.; Pereira, A.C.; Pappa, G.L. A customized classification algorithm for credit card fraud detection. *Eng. Appl. Artif. Intell.* **2018**, *72*, 21–29. [CrossRef]
32. Robinson, W.N.; Aria, A. Sequential fraud detection for prepaid cards using hidden Markov model divergence. *Expert Syst. Appl.* **2018**, *91*, 235–251. [CrossRef]
33. Hajek, P.; Henriques, R. Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods. *Knowl.-Based Syst.* **2017**, *128*, 139–152. [CrossRef]
34. Craja, P.; Kim, A.; Lessmann, S. Deep learning for detecting financial statement fraud. *Decis. Support Syst.* **2020**, *139*, 113421. [CrossRef]
35. Ravisankar, P.; Ravi, V.; Rao, G.R.; Bose, I. Detection of fi nancial statement fraud and feature selection using data mining techniques. *Decis. Support Syst.* **2011**, *50*, 491–500. [CrossRef]
36. Gao, Y.; Sun, C.; Li, R.; Li, Q.; Cui, L.; Gong, B. An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services. *IEEE Access* **2018**, *6*, 60059–60068. [CrossRef]
37. Huang, S.-Y.; Tsaih, R.-H.; Yu, F. Topological pattern discovery and feature extraction for fraudulent financial reporting. *Expert Syst. Appl.* **2014**, *41*, 4360–4372. [CrossRef]
38. Peng, J.; Li, Q.; Li, H.; Liu, L.; Yan, Z.; Zhang, S. Fraud Detection of Medical Insurance Employing Outlier Analysis. In Proceedings of the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanjing, China, 9–11 May 2018; pp. 341–346.
39. van Capelleveen, G.; Poel, M.; Mueller, R.M.; Thornton, D.; van Hillegersberg, J. Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *Int. J. Account. Inf. Syst.* **2016**, *21*, 18–31. [CrossRef]

40. Anbarasi, M.S.; Dhivya, S. Fraud detection using outlier predictor in health insurance data. In Proceedings of the 2017 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 23–24 February 2017; pp. 1–6.
41. Sundarkumar, G.G.; Ravi, V.; Siddeshwar, V. One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 10–12 December 2015; pp. 1–7. [CrossRef]
42. Subudhi, S.; Panigrahi, S. Effect of Class Imbalanceness in Detecting Automobile Insurance Fraud. In Proceedings of the 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), ChangSha, China, 21–23 September 2018; pp. 528–531.
43. Fayyomi, M.; Eleyan, D.; Eleyan, A. A Survey Paper On Credit Card Fraud Detection Techniques. *Int. J. Adv. Res. Comput. Eng. Technol.* **2021**, *3*, 827–832.
44. Wang, Y.; Xu, W. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decis. Support Syst.* **2018**, *105*, 87–95. [CrossRef]
45. Gepp, A.; Kumar, K.; Bhattacharya, S. Lifting the numbers game: Identifying key input variables and a best-performing model to detect financial statement fraud. *Account. Financ.* **2021**, *61*, 4601–4638. [CrossRef]
46. Perols, L.; Lougee, B.A. The relation between earnings management and financial statement fraud. *Adv. Account.* **2011**, *27*, 39–53. [CrossRef]
47. Wang, Q.; Xu, W.; Huang, X.; Yang, K. Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. *Neurocomputing* **2019**, *347*, 46–58. [CrossRef]
48. Islam, S.R.; Ghafoor, S.K.; Eberle, W. Mining Illegal Insider Trading of Stocks: A Proactive Approach. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1397–1406. [CrossRef]
49. Kulkarni, P.M.; Domeniconi, C. Network-based anomaly detection for insider trading. *arXiv* **2017**, arXiv:1702.05809.
50. Mirtaheri, M.; Abu-El-Haija, S.; Morstatter, F.; Steeg, G.V.; Galstyan, A. Identifying and Analyzing Cryptocurrency Manipulations in Social Media. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 607–617. [CrossRef]
51. Monamo, P.M.; Marivate, V.; Twala, B. A Multifaceted Approach to Bitcoin Fraud Detection: Global and Local Outliers. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 188–194. [CrossRef]
52. Vasek, M.; Moore, T. There's No Free Lunch, Even Using Bitcoin: Tracking the Popularity and Profits of Virtual Currency Scams BT–Financial Cryptography and Data Security. In Proceedings of the International Conference on Financial Cryptography and Data Security, Kota Kinabalu, Malaysia, 1–5 March 2015; pp. 44–61.
53. Monamo, P.; Marivate, V.; Twala, B. Unsupervised learning for robust Bitcoin fraud detection. In Proceedings of the 2016 Information Security for South Africa (ISSA), Johannesburg, South Africa, 17–18 August 2016; pp. 129–134. [CrossRef]
54. Li, X.; Ying, S. Lib-SVMs Detection Model of Regulating-Profits Financial Statement Fraud Using Data of Chinese Listed Companies. In Proceedings of the 2010 International Conference on E-Product E-Service and E-Entertainment, Henan, China, 7–9 November 2010; pp. 1–4. [CrossRef]
55. Throckmorton, C.S.; Mayew, W.J.; Venkatachalam, M.; Collins, L.M. Financial fraud detection using vocal, linguistic and fi nancial cues. *Decis. Support Syst.* **2015**, *74*, 78–87. [CrossRef]
56. Glancy, F.H.; Yadav, S.B. A computational model for fi nancial reporting fraud detection. *Decis. Support Syst.* **2011**, *50*, 595–601. [CrossRef]
57. Mareeswari, V.; Gunasekaran, G. Prevention of credit card fraud detection based on HSVM. In Proceedings of the 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 25–26 February 2016; pp. 1–4.
58. Humpherys, S.L.; Mof, K.C.; Burns, M.B.; Burgoon, J.K.; Felix, W.F. Identi fi cation of fraudulent fi nancial statements using linguistic credibility analysis. *Decis. Support Syst.* **2011**, *50*, 585–594. [CrossRef]
59. Li, X.; Xu, W.; Tian, X. How to protect investors? A GA-based DWD approach for financial statement fraud detection. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; pp. 3548–3554. [CrossRef]
60. Karlos, S.; Fazakis, N.; Kotsiantis, S.; Sgarbas, K. Semi-supervised forecasting of fraudulent financial statements. In Proceedings of the 20th Pan-Hellenic Conference on Informatics, Patras, Greece, 10–12 November 2016. [CrossRef]
61. Özçelik, M.H.; Duman, E.; Işik, M.; Çevik, T. Improving a credit card fraud detection system using genetic algorithm. In Proceedings of the 2010 International Conference on Networking and Information Technology, Manila, Philippines, 11–12 June 2010; pp. 436–440.
62. Rizki, A.; Surjandari, I.; Wayasti, R.A. Data mining application to detect financial fraud in Indonesia's public companies. In Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech), Bandung, Indonesia, 25–26 October 2017; pp. 206–211.
63. Chen, S. Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus* **2016**, *5*, 1–16. [CrossRef] [PubMed]
64. Yao, J.; Zhang, J.; Wang, L. A financial statement fraud detection model based on hybrid data mining methods. In Proceedings of the 2018 international conference on artificial intelligence and big data (ICAIBD), Chengdu, China, 26–28 May 2018; pp. 57–61. [CrossRef]

65. Rajak, I.; Mathai, K.J. Intelligent fraudulent detection system based SVM and optimized by danger theory. In Proceedings of the 2015 International Conference on Computer, Communication and Control (IC4), Indore, India, 10–12 September 2015; pp. 1–4. [CrossRef]
66. Jeragh, M.; Alsulaimi, M. Combining Auto Encoders and One Class Support Vectors Machine for Fraudulant Credit Card Transactions Detection. In Proceedings of the 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 30–31 October 2018; pp. 178–184. [CrossRef]
67. Kho, J.R.D.; Vea, L.A. Credit card fraud detection based on transaction behavior. In Proceedings of the TENCON 2017-2017 IEEE Region 10 Conference, Penang, Malaysia, 5–8 November 2017; pp. 1880–1884. [CrossRef]
68. Behera, T.K.; Panigrahi, S. Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network. In Proceedings of the 2015 Second International Conference on Advances in Computing and Communication Engineering, Dehradun, India, 1–2 May 2015; pp. 494–499.
69. HaratiNik, M.R.; Akrami, M.; Khadivi, S.; Shajari, M. FUZZGY: A hybrid model for credit card fraud detection. In Proceedings of the 6th International Symposium on Telecommunications (IST), Tehran, Iran, 6–8 November 2012; pp. 1088–1093.
70. Malini, N.; Pushpa, M. Analysis on credit card fraud identification techniques based on KNN and outlier detection. In Proceedings of the 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB), Chennai, India, 27–28 February 2017; pp. 255–258. [CrossRef]
71. Benchaji, I.; Douzi, S.; ElOuahidi, B. Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection. In Proceedings of the International Conference on Advanced Information Technology, Services and Systems, Mohammedia, Morocco, 17–18 October 2018; pp. 1–5. [CrossRef]
72. Case, B. Recognizing Debit Card Fraud Transaction Using CHAID and K-Nearest Neighbor: Indonesian Bank case. In Proceedings of the 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Yogyakarta, Indonesia, 10–12 November 2016.
73. Bhusari, V.; Patil, S. Study of Hidden Markov Model in credit card fraudulent detection. In Proceedings of the 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, India, 29 February–1 March 2016; pp. 1–4.
74. Sahin, Y.; Bulkan, S.; Duman, E. A cost-sensitive decision tree approach for fraud detection. *Expert Syst. Appl.* **2013**, *40*, 5916–5923. [CrossRef]
75. Duman, E.; Ozcelik, M.H. Detecting credit card fraud by genetic algorithm and scatter search. *Expert Syst. Appl.* **2011**, *38*, 13057–13063. [CrossRef]
76. Sahin, Y.; Duman, E. Detecting credit card fraud by ANN and logistic regression. In Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, 15–18 June 2011; pp. 315–319.
77. Ghobadi, F.; Rohani, M. Cost sensitive modeling of credit card fraud using neural network strategy. In Proceedings of the 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, Iran, 14–15 December 2016; pp. 1–5.
78. Awoyemi, J.O.; Adetunmbi, A.O.; Oluwadare, S.A. Credit card fraud detection using machine learning techniques: A comparative analysis. In Proceedings of the 2017 international conference on computing networking and informatics (ICCNI), Ota, Nigeria, 29–31 October 2017; pp. 1–9. [CrossRef]
79. Mishra, A.; Ghorpade, C. Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques. In Proceedings of the 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 24–25 February 2018; pp. 1–5. [CrossRef]
80. Kirlidog, M.; Asuk, C. A Fraud Detection Approach with Data Mining in Health Insurance. *Procedia-Soc. Behav. Sci.* **2012**, *62*, 989–994. [CrossRef]
81. Peng, H.; You, M. The Health Care Fraud Detection Using the Pharmacopoeia Spectrum Tree and Neural Network Analytic Contribution Hierarchy Process. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August 2016; pp. 2006–2011. [CrossRef]
82. Bauder, R.; da Rosa, R.; Khoshgoftaar, T. Identifying Medicare Provider Fraud with Unsupervised Machine Learning. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 7–9 July 2018; pp. 285–292.
83. Bauder, R.A.; Khoshgoftaar, T.M.; Richter, A.; Herland, M. Predicting Medical Provider Specialties to Detect Anomalous Insurance Claims. In Proceedings of the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), San Jose, CA, USA, 6–8 November 2016; pp. 784–790.
84. Badriyah, T.; Rahmaniah, L.; Syarif, I. Nearest Neighbour and Statistics Method based for Detecting Fraud in Auto Insurance. In Proceedings of the 2018 International Conference on Applied Engineering (ICAE), Batam, Indonesia, 3–4 October 2018; pp. 1–5. [CrossRef]
85. Zhou, Y.; Wang, X.; Zhang, J.; Zhang, P.; Liu, L.; Jin, H.; Jin, H. Analyzing and Detecting Money-Laundering Accounts in Online Social Networks. *IEEE Netw.* **2017**, *32*, 115–121. [CrossRef]
86. Mhamane, S.S.; Lobo, L.M.R.J. Internet banking fraud detection using HMM. In Proceedings of the 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Karur, India, 26–28 July 2012; pp. 1–4.
87. Faraji, Z.; States, U. A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study. *J. Manag.* **2022**, *5*, 49–59. [CrossRef]

88. Bhavitha, B.K.; Rodrigues, A.P.; Chiplunkar, N.N. Comparative study of machine learning techniques in sentimental analysis. In Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 10–11 March 2017; pp. 216–221. [CrossRef]
89. Carta, S.; Fenu, G.; Recupero, D.R.; Saia, R. Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. *J. Inf. Secur. Appl.* **2019**, *46*, 13–22. [CrossRef]
90. Rb, A.; Kr, S.K. Credit card fraud detection using artificial neural network. *Glob. Transit. Proc.* **2021**, *2*, 35–41. [CrossRef]
91. Pradeep, G.; Ravi, V.; Nandan, K.; Deekshatulu, B.L.; Bose, I.; Aditya, A. Fraud Detection in Financial Statements Using Evolutionary Computation Based Rule Miners. In Proceedings of the International Conference on Swarm, Evolutionary, and Memetic Computing, Hyderabad, India, 18–19 December 2015; pp. 239–250. [CrossRef]
92. Hajek, P. Interpretable Fuzzy Rule-Based Systems for Detecting Financial Statement Fraud. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Crete, Greece, 25–27 June 2019; pp. 1–12.
93. Khan, A.; Singh, T.; Sinhal, A.; Khan, A.; Singh, T. Implement credit card fraudulent detection system using observation probabilistic in hidden Markov model. In Proceedings of the 2012 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, 6–8 December 2012; pp. 1–6. [CrossRef]
94. Wang, X.; Wu, H.; Yi, Z. Research on Bank Anti-Fraud Model Based on K-Means and Hidden Markov Model. In Proceedings of the 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 780–784. [CrossRef]
95. Song, R.; Huang, L.; Cui, W.; Vanthienen, J. Fraud Detection of Bulk Cargo Theft in Port Using Bayesian Network Models. *Appl. Sci.* **2020**, *10*, 1056. [CrossRef]
96. Dang, T.K.; Tran, T.C.; Tuan, L.M. Machine Learning Based on Resampling Approaches and Deep Reinforcement Learning for Credit Card Fraud Detection Systems. *Appl. Sci.* **2021**, *11*, 10004. [CrossRef]
97. Bouchti, E.; Chakroun, A.; Abbar, H.; Okar, C. Fraud detection in banking using deep reinforcement learning. In Proceedings of the 2017 Seventh International Conference on Innovative Computing Technology (INTECH), Luton, UK, 16–18 August 2017; pp. 58–63.
98. Zouboulidis, E.; Kotsiantis, S. Forecasting fraudulent financial statements with committee of cost-sensitive decision tree classifiers. In *Hellenic Conference on Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 57–64.
99. Hassanzadeh, R. A Nomaly Detection in Online Social Networks: Using Data-Mining Techniques and Fuzzy. Ph.D. Thesis, Queensland University of Technology, Brisbane City, QLD, Australia, 2014.
100. Shah, V.; Shah, P.; Shetty, H.; Mistry, K. Review of Credit Card Fraud Detection Techniques. In Proceedings of the 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 29–30 March 2019; pp. 1–7. [CrossRef]
101. Deng, Q. Detection of fraudulent financial statements based on Naïve Bayes classifier. In Proceedings of the 2010 5th International Conference on Computer Science & Education, Hefei, China, 24–27 August 2010; pp. 1032–1035. [CrossRef]
102. Ahmed, M.; Mahmood, A.N.; Islam, R. A survey of anomaly detection techniques in financial domain. *Futur. Gener. Comput. Syst.* **2016**, *55*, 278–288. [CrossRef]
103. Uchhana, N.; Ranjan, R.; Sharma, S.; Agrawal, D.; Punde, A. Literature Review of Different Machine Learning Algorithms for Credit Card Fraud Detection. *Int. J. Innov. Technol. Explor. Eng.* **2021**, *10*, 101–108. [CrossRef]
104. Abbasi, A.; Albrecht, C.; Vance, A.; Hansen, J. Metafraud: A meta-learning framework for detecting financial fraud. *Mis Q.* **2012**, *36*, 1293–1327. [CrossRef]
105. Moepya, S.O.; Nelwamondo, F.V.; Twala, B. Increasing the detection of minority class instances in financial statement fraud. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Kanazawa, Japan, 3–5 April 2017; Volume 2, p. 2017.
106. Chen, S.; Goo, Y.-J.J.; Shen, Z.-D. A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements. *Sci. World J.* **2014**, *2014*, 1–9. [CrossRef] [PubMed]
107. Patel, H.; Parikh, S.; Patel, A.; Parikh, A. An Application of Ensemble Random Forest Classifier for Detecting Financial Statement Manipulation of Indian Listed Companies. In *Recent Developments in Machine Learning and Data Analytics*; Springer: Singapore, 2019.
108. Hobson, L.; Mayew, W.J. Analyzing Speech to Detect Financial Misreporting Analyzing Speech to Detect Financial Misreporting. *J. Account. Res.* **2010**, *2*, 349–392.
109. Li, Y.; Yan, C.; Liu, W.; Li, M. Research and application of random forest model in mining automobile insurance fraud. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 1756–1761. [CrossRef]
110. Kowshalya, G.; Nandhini, M. Predicting Fraudulent Claims in Automobile Insurance. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 1338–1343.
111. Bauder, R.; Khoshgoftaar, T. Medicare Fraud Detection Using Random Forest with Class Imbalanced Big Data. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 7–9 July 2018; pp. 80–87.

112. Li, S.-H.; Yen, D.C.; Lu, W.-H.; Wang, C. Identifying the signs of fraudulent accounts using data mining techniques. *Comput. Hum. Behav.* **2012**, *28*, 1002–1013. [CrossRef]
113. Bartoletti, M.; Pes, B.; Serusi, S. Data Mining for Detecting Bitcoin Ponzi Schemes. In Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 20–22 June 2018; pp. 75–84. [CrossRef]
114. Zhang, W.; He, X. An Anomaly Detection Method for Medicare Fraud Detection. In Proceedings of the 2017 IEEE International Conference on Big Knowledge (ICBK), Hefei, China, 9–10 August 2017; pp. 309–314.
115. Deng, Q.; Mei, G. Combining self-organizing map and K-means clustering for detecting fraudulent financial statements. In Proceedings of the 2009 IEEE International Conference on Granular Computing, Nanchang, China, 17–19 August 2009; pp. 126–131. [CrossRef]
116. Sael, N.; Benabbou, F. ScienceDirect ScienceDirect Performance of machine learning techniques in the detection of Performance of machine learning techniques in the detection of financial frauds financial frauds. *Procedia Comput. Sci.* **2018**, *148*, 45–54.
117. Liang, J.; Lv, W. Research on detecting technique of financial statement fraud based on Fuzzy Genetic Algorithms BPN. In Proceedings of the 2009 International Conference on Management Science and Engineering, Nanchang, China, 17–19 August 2009; pp. 1462–1468. [CrossRef]
118. Xiaoyun, W.; Danyue, L. Hybrid outlier mining algorithm based evaluation of client moral risk in insurance company. In Proceedings of the 2010 2nd IEEE International Conference on Information Management and Engineering, Chongqing, China, 17–19 September 2010; pp. 585–589. [CrossRef]
119. Bauder, R.A.; Khoshgoftaar, T.M. Medicare Fraud Detection Using Machine Learning Methods. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 858–865.
120. Pejic-bach, M. Invited Paper: Profiling Intelligent Systems Applications in Fraud Detection and Prevention: Survey of Research Articles Profiling intelligent systems applications in fraud detection and prevention: Survey of research articles. In Proceedings of the 2010 International Conference on Intelligent Systems, Modelling and Simulation, Liverpool, UK, 27–29 January 2010.
121. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
122. D'Addio, R.M.; Manzato, M.G. A Collaborative Filtering Approach Based on User's Reviews. In Proceedings of the 2014 Brazilian Conference on Intelligent Systems, Washington, DC, USA, 18–22 October 2014; pp. 204–209. [CrossRef]
123. Paruchuri, H. Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review. *ABC J. Adv. Res.* **2017**, *6*, 113–120. [CrossRef]
124. Silva, B.; Marques, N.; Panosso, G. Applying neural networks for concept drift detection in financial markets. *CEUR Workshop Proc.* **2012**, *960*, 43–47.