State-of-The-Art in Big Data Application Techniques to Financial Crime : A Survey

Abiodun Esther Omolara[†], Aman Jantan, Oludare Isaac Abiodun[†], Manmeet Mahinderjit Singh[†], Mohammed Anbar[†], Kemi Victoria Dada[††]

[†]School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
Correspondent Author: Aman Jantan
[††]Department of Statistics, Faculty of Science, Ahmadu Bello University Zaria, Nigeria

Summary
The goal is to survey big data techniques applications to financial crime prevention and detection in more than two decades. Also, to determine the industrial sector of financial crime that has gathered the most interest and which fields still lack research. The study describes the most common methods for detecting financial crime, indicating some of the current research problems, trends, and issues in big data application. Meanwhile, the survey focuses on financial crime detection techniques applied in several sectors such as banks, computer networks, insurance, securities, exchange commodities, stock markets and money laundering. The methods considered in the survey include; big data analytics foundational technologies and big data analytics emerging research. In most of the data-collection strategies and data analysis, there was a shift from traditional data collection method to computer-based methods. In the aspect of data analysis, there was an increase in the use of descriptive, inferential statistical analysis. We made an evaluation of the detecting techniques based on data analysis factors such as processing speed, latency, volume, performance, fault tolerance, scalability, and accuracy. Then we propose that anomaly, data-mining, clustering, hybrid-technologies, neural networks, rough clustering, k-Means clustering, neuro-fuzzy, genetic algorithms and fuzzy support vector machine models are performing better than other methods currently in practice. However, more research is required in big data infrastructures like hardware equipment, software licensing, and maintenance which are still very expensive. Likewise, further research is needed in the human analysis of big data approaches to financial crime detection because of the challenges of sorting out information.

Key words:
Financial-crime, big data techniques, detecting, prevention, methods.

1. Introduction

The expeditious and stupendous increasing rate in applied soft computing worldwide has a major contribution toward the total output of big data application to financial crime detection. According to Tiwari, Knowles, Avineri, Dahal, and Roy, [1] “Soft Computing is a complex of methodologies that adopt reasoning, uncertainty, imprecision, and partial truth to mimic human capability of making decisions in real-life, and ambiguous environments”. The Big data analysis has potential value for enhancing decision-making process, which recently attracts increasing interest from both academics and practitioners. Big Data (BD) refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze [2]. BD can be found in every business function and industry and are a major factor for production. Examples of big data used cases are fraud detection, log analytics, sentiment analysis and social media, energy sector, risk modeling, and management.

Although, the Big data can generate financial value across sectors; global personal location data, healthcare, manufacturing, public sector administration, retail etc. However, there are issues to be addressed in BD development phase to improve performance these includes; data policies like analytical software, storage, computing, and new types of analyses. Others include technology and techniques issues such as security, privacy, intellectual property, and liability. An issue of access to data like integrated multiple data sources and industrial structure such as absence of competitive pressure in public domain.

In recent years, big data techniques (BDTs) have become a known and useful tool for prediction, and pattern recognition or classification in many fields. Big data techniques wide applications can be explained in terms of data analysis factors like processing speed, latency, performance, fault tolerance, volume, scalability, and accuracy. These factors give more reason that BDTs are effective and efficient in providing a high level of capability in handling various complex and non-complex problems in many spheres of life such as in finance, management, security, engineering, agriculture, medical science, education, and science. Others include banking, insurance, properties management, manufacturing, transportation, computer security, marketing, energy, and problems that are beyond the computational capability of traditional procedures and classical mathematics.

Manuscript received July 5, 2018
Manuscript revised July 20, 2018

Despite the useful tool of BDT's in prediction, pattern recognition and efficiency of its application. However, there is need to address some problems in the application such as high costs of Big Data's infrastructures, and maintenance [3,4]. The infrastructures like hardware equipment and software licensing is very expensive, as well as maintenance is expensive because of the high level of technical skill. Also, sorting through data for valuable information is tedious and challenging in terms of human analysis [3]. As reported by Akerkar [5] and Zicari [2], the main challenges of BD can be grouped into three categories, based on the data lifecycle: data, process and management issues as in Figure 1.

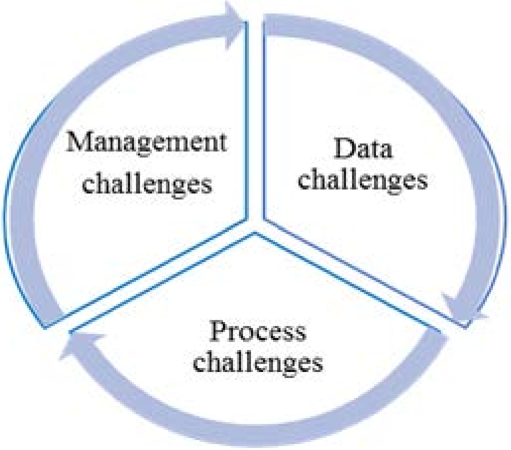

Fig. 1 Data lifecycle: data, process and management issues

The Fig. 1 illustrates the data lifecycle starting from raw data to processing and to management.
(i) Data challenges relate to the characteristics of the data itself, an example, the data challenges include data volume, variety, velocity, veracity, volatility, quality, discovery, and dogmatism.
(ii) Process challenges are related to series of how techniques: how to capture data, how to integrate data, how to transform data, how to select the right model for analysis and how to provide the results.
(iii) Management challenges cover for example privacy, security, governance and ethical aspects.
More research is needed into the big data environment supports to resolve the issues of cybersecurity in terms of finding the attacker. However, in this study, the challenges faced in applying BD by an analyst are explained in areas of the financial fraud detection, network forensics, data privacy issues and data provenance problems.
Detection and prevention of financial crime is a method of reducing financial crime incidents and identifying early risks and issues by implementing effective and efficient operational controls. The focus is on enterprise crime and misuse management using advanced analytics and integrated technology platforms.
However, some of the primary issues and challenges associated with the current practices and potential future direction of research have been identified which are required to be addressed before new fraud techniques can be adopted. Therefore, this section presents some of the primary issues associated with inancial crime detection and suggest areas for future research.

2. Financial Fraud Detection: Current research issues in financial crime and future direction.

There are still aspects of intelligent financial fraud detection that are yet studying. The current research issues in financial crime include;

- The theory that explains the definitions, or predicts, a set of interrelated constructs, and propositions that present a systematic view of phenomena by specifying relations among variables, with the aim of explaining natural phenomena. The systematic view might be a discussion, a rationale, or an argument, and it helps to explain or predict phenomena that occur in the real world. Behavioral theories of financial crime are theories that are developed explicitly to explain individual aspects of financial crime. Thus, theory acts as an educational device that creates insights into criminal phenomena. Hansen, [6] suggests that distinction can be made between economic, business and elite crimes. Cain, [6] argues that the general piece of evidence available is that criminals are more responsive to changes in the chance of being caught than to changes in the consequence.

- Classification problems: Big data and computational intelligence and based financial fraud detection such as feature selection, analysis of the problem domain, and parameter tuning need to be addressed.
- Disproportionate misclassification costs: Financial fraud detection is basically a classification problem with a difference in misclassification costs. An investigation on the performance of detection techniques with respect to misclassification costs is an area that requires further attention.
- Generic framework: Since there are many classifications or kinds of financial fraud, a generic framework which can be applied to multiple financial fraud categories would be valuable to address the problem [7].

- Fraud types and detection methods: Financial crime is a verse field and there has been a much inequality in both detection methods studied and fraud types: Some have been studied widely while others, such as hybrid techniques, have not been well studied which require being study.
- Privacy issues: notably financial fraud is a sensitive topic and most stakeholders are reluctant to share information on the subject. This has brought about experimental issues such as undersampling techniques.
- Computational performance: As a sensitive and high-cost problem it is necessary for financial fraud to be detected immediately. Currently, few researches have been conducted on the computational intelligence performance of fraud detection approaches for use in real-time situation or environment.
- Evolving problem: Financial fraudsters are continually changing tactics to remain undetected. As such detection methods are required constantly to counter new fraud tactics.

3. Other Challenges and Future Directions

Other current challenges in financial crime detection include the regulatory change which has been a major challenge for most financial institutions [8]. The regulatory landscape impacts many different business areas, including those involved in fighting financial crime such as anti-fraud management, anti-money laundering prevention, and cybersecurity. Identifying exactly which regulation impacts a specific business area is a hugely complex task. This multi-dimensional complexity impacts people, technology, and processes in a way that the current silo-based approach used by most financial institutions in combating financial crime is no longer sufficient. Another main challenge is ensuring compliance while protecting the institution from cyber threats is becoming more challenging. Both require a new compliance and security approach.

There are plethora factors enveloping that are making the fight against financial crime a major and costly challenge. Firstly, regulators across the world are trying to solve similar issues. However, they all tackling them in a unique way, with different rules. This means that, for national, regional or global banks with operations in several jurisdictions, the institution has to determine which rules apply to each transaction. In fact the larger the institution and the more complex its operations, the more complex compliance becomes [8].

Secondly, the development of instantaneous electronic payment systems across multiple devices and the globalization of finance have had an unparalleled impact on the evolution of fraud is a serious challenge. Hopefully, by the end of 2018, many countries will process transactions in real time such as using a blockchain, distributed database protocol, and permissionless. Hence, delinquent behavioral patterns will be detected within nanoseconds. Both trends also present considerable challenges for customer due diligence, sanctions screening, and transaction monitoring.

This means that in the banking world, instead of checking and stopping transactions that are processed overnight, or in a batch, everything now needs to be done in real time. Also, if an alert is raised, it needs to be checked immediately. Then this will raise the cost of processing each alert. Recent analyses estimate that a false positive alert costs a bank more than $30 [8]. With global banks processing millions of transactions and most dealing with false positive rates exceeding 5% of all alerts, the costs of managing alerts may become unsustainable.

Thirdly, the threat of financial crime now includes financial fraud, bribery, and money laundering, as well as a raft of new cybercrimes, which can be even more difficult to monitor and detect or prevent. The threat challenge has become so complex that it is impossible for individual units or departments operating in silos to effectively control or manage it. Fourthly, recent research has shown that protection of client assets is the highest priority issue for corporates and consumers when selecting a financial provider [8]. The threat of financial crime runs the gamut of financial fraud and abuse such as money laundering, bribery, and cybercrime. It has, therefore, become too broad to be handled by established divisions or departments.

The human motivation aroused from the alarmingly increasing rate in the number of computer-based financial crime records in the last decade. Firstly, the past and present financial crime in the banking sector, insurance industry, computer networks, securities, and exchange commodities markets and money laundering phenomena which has resulted to loss of billions of dollars in the economy make motivation of this research for possible detecting solution. Secondly, the fact that many nations are seeking a solution to financial crime has called for research interest into the survey on financial crime detection using big data approach.

4. Background

With the increasing number of high-profile of financial crime, cyber-attacks, and the increasingly documented connection with the financial crime, there is an urgent need to address the menace. Financial fraud can be defined as a way of stealing money or property or both for self-gain or illegally used of them for the benefit. The government, financial institution, corporate organizations, and consumers lose billions of dollars due to fraud anually and fraudsters continuously try to use their method or find new gambit to commit illegal actions [10,11].

Recent increasing dependence on new technologies such as automated teller machine (ATM), online payment, mobile computing and cloud computing has made the problem complex. The techniques for hiding proceeds of financial crime include transferring the money from the country, buying businesses by which funds can be channeled, purchasing an easy transportation valuable, transfer pricing, and using “underground banks.

Though, financial institutions are slowly waking up to the new ways of fighting fraud [12]. Most banks detect financial fraud or crime by cash transaction monitoring, identify cash transactions just below regulatory reporting thresholds billing, identify an unusually large number of the waived fee by branch or by the employee.

Though, in recent time whistle-blower policy in some countries have attracted people’s attention because it is a financial detection technique whereby an anonymous person secretly contacts government officers about stolen money availability to receive monetary reward offered. Some former and present employers have stolen so much money that is meant for the public. This whistle-blower policy fraud detection method that most countries adopted has attracted more than $10 million to multiple whistle-blowers annually.

5. Financial frauds classification

Financial frauds can be classified into; bank fraud, insurance fraud, computer fraud, securities and exchange commodities fraud, other related financial fraud. Theses classification can briefly explain as follow;

(i) Bank fraud: is executed to defraud a financial institution; or to obtain fund, money, assets, credits, securities, or other property owned by a financial institution, using fraudulent pretends.” Such as mortgage fraud, money laundering, and so on.

(ii) Insurance fraud: is the ones which occur in between the insurance process. It can happen while in the application, billing, rating, claims, eligibility process and are dedicated mostly to healthcare providers, consumers, agents or brokers, company employees and others.

(iii) Security and exchange commodities fraud: According to definition by CULS, security frauds include theft from manipulation of the market, securities accounts and wire fraud [18]. It widely provides market manipulation, high yield investment fraud, commodities fraud, foreign exchange fraud, late-day trading, broker embezzlement, etc.

(iv) Computer-based financial fraud: is a cybercrime in which computer and electronics machine is used as a medium to steal money or property. An example is the use of computer network system such as ATM, internet, and credit card to hack people’s financial account details to make a withdraw of money for personal gain.

(v) Other related financial fraud: these are those which includes classes above, such that mass marketing and corporate fraud.

The framework for financial fraud classification in the survey is depicted in Figure 2.

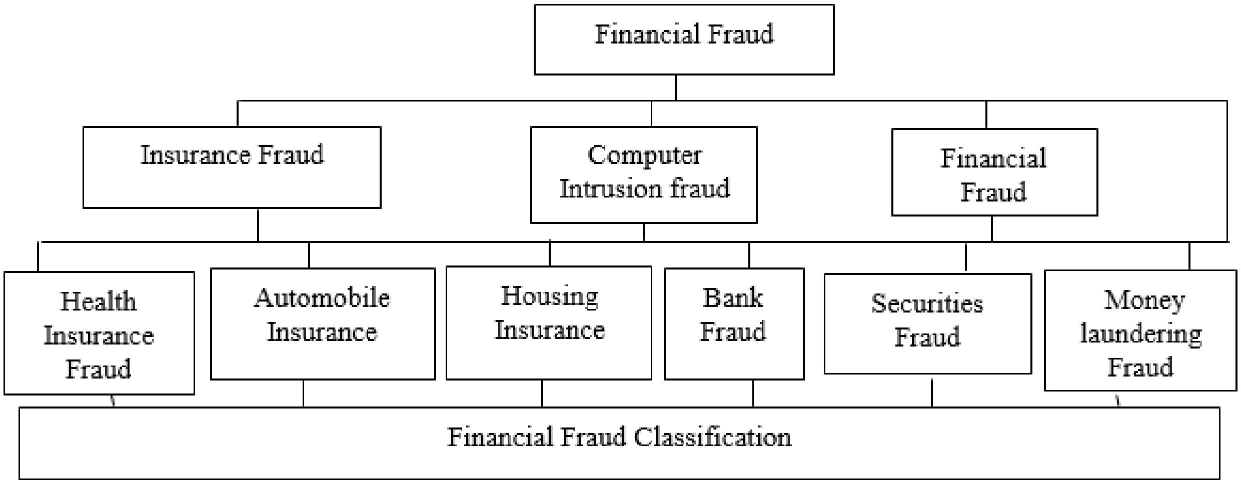

Fig. 2 Review Framework for Financial Fraud Classification

6. Economic analysis of financial crime

Financial crime has a corrosive effect on a country's *economy*, government, and social well-being of the people. The negative impact of financial crimes includes; slow-growing economy, poverty among the populace, unemployment problem and many social ills.

Unfortunately, those people that commit financial crimes for selfish gain want not only to escape punishment, but they want to enjoy the fruits of their ill-gotten wealth. However, this enjoyment usually takes the form of immediate and conspicuous consumption [19]. Notably, financial crime has no boundary, as it is common in developing countries, so also the crime is committed in a developed economy [19,20,21,22].

In another perspective financial crimes referred to a broad category of crimes against property, committed by individuals and organizations to obtain a personal or business advantage [23]. An example include bank fraud, insurance fraud, computer crime, money laundering, credit/debit card fraud, embezzlement, counterfeiting, mortgage fraud, and insider trading, and so on. These crimes cost several billions of dollars a year and affect the lives of millions of people. The recent announcement that Equifax security leak people’s account might have affected 145.5 million American consumers [9]. As consumers account, information was accessed by hackers between mid-May and July this year. The recent news from one of the big four credit reporting agencies sent shock waves throughout the nation.

The stolen detail includes names, addresses, social security numbers, dates of birth and, driver’s license numbers, credit card and credit dispute information. This stolen information makes financial crime easy for the fraudsters, because of this Equifax attacked the stock market loss value was more than $4 billion and an insured loss of $125 million from a massive data breach by Equifax Inc [9]. Equifax is a global information solutions film that uses unique data, technology, innovative analytics, and industry expertise to power establishment and individuals around the world. By transforming knowledge into insights that help more informed business and personal decisions. Equifax headquarters is in Atlanta, Ga., Equifax operates or has investments in 24 countries in part of America, Europe, Central and South America, and Asia Pacific region.

Additionally, and in another angle financial crimes may involve criminal acts, such as armed robbery, elder abuse, burglary, and even violent crime such as or murder or robbery. Financial crimes may be carried out by individuals, corporations, or by organized crime groups. Victims may include individuals, a group of people, organizations, institution, government at all levels, regional economies, and entire global economies. While “fraud” may define as a wrongful or criminal deception intended to result in a financial, material or personal gain.

Security and financial risk managers are responsible for online fraud detection report that mobile interfaces, retail locations and call centers, payment systems are all vulnerable to financial fraud attacks. The expansion and reach of financial fraud today creates losses so extensive that observers have a difficult time measuring cost [24]. As electronic fraud attacks become more sophisticated and complicated, it is time to rethink functional online fraud detection and protection architecture [24].

7. Bank fraud destroys customer confidence

Banking fraud referred to the use of illegal means to get money, assets, or other landed property held or owned by a financial institution, or to extort money from customers or depositors by fraudulently assuming the position as a bank or representing other financial institution. Bank fraud destroys customer confidence because the money may no longer be safe as the most bank have suddenly distressed overnight, and the customer’s monies lost. The most common financial fraud in most countries is bank frauds and government agencies fraud. In recent times bank fraud includes illegal fund transfer and withdrawals; Posting of fictitious credits; use of unauthorized overdraft; conversion of banks money into personal use; Presentation of forged cheques; and granting of unauthorized loans. Also, abuse of medical scheme; ghost workers fraud resulting in millions of dollars paid into private pockets; Illegal conversion of pension funds in various agencies and ministries; insider abuse and over-invoicing crimes [25].

Banks have the option of doing fraud detection either in real time or once every twenty-four hours because personal accounts are the responsibility of the banks, fraud detection for private accounts is usually carried out in real time. Similar to banking fraud is a mortgage fraud with different illegal schemes related to the misrepresentation or misstatement of mortgage documents to defraud such as a lender or a homeownership [26,27,28, 29,30,31,32].

Most current systems used to detect suspicious activity in banks are based on rules. While these systems are fast and pick up most known issues, new crimes require constant review and regular updating of the rules to stay one step ahead [15]. Artificial intelligence, on the other hand, can “learn” and provide valuable data that can be used to enhance protection [15].

8. Money laundering destroys value

Money laundering can be described as a crime involving an illicit movement of stolen money for gains in legitimate channels to disguise the money's as lawful source [33]. Money laundering destroys money value because a huge amount of money is taken away from the nation without been use to develop the economy. Again, money laundering activities destroy money value because its facilitate economic crimes like corruption, tax evasion, terrorism, drug and human trafficking by transferring or holding the funds necessary to commit these crimes. It can be detrimental to an organization’s reputation and its bottom line [33]. Regrettably, global money laundering transactions estimated between 2% to 5% of global GDP, or the roughly U.S. $1-2 trillion annually.

9. Securities and exchange commodities fraud

Securities fraud is a crime in which a corporate officer, for example, discloses confidential information related to its stock or makes misleading statements about the company’s stock performance. Similarly, securities fraud broadly refers to deceptive practices relating to the offering for sale of securities [33,34]. A commodities fraud includes the sale or purported sale of a commodity, i.e., raw materials or semi-finished goods that sold on an exchange such as gold or coffee, by illegal way. There are two common types of commodities investment frauds these are;

(i) Foreign currency exchange (forex) fraud - perpetrators of forex frauds may entice people into investing spot foreign currency market by false claims and high-pressure sales tactics.

(ii) Precious metals fraud - These schemes offer investment opportunities in metals commodities such as rare earth, silver and, gold.

10. Insurance fraud

Insurance Fraud Insurance frauds involve an insurance company, agent or other person being deceived by individuals to achieve monetary gains to which they are not entitled. Again, insurance fraud is said to have happened when someone has put false information on an insurance application and when misleading or incorrect information given, or crucial information omitted in an insurance claim or transaction. In recent years, insurance financial crime cases are on the rise. According to insurance industry statistics, the amount recorded of insurance fraud has accounted for 10% to 30% of the total payment in an insurance firm, for some special insurance, it's has grown to 50% [32]. There are multiple areas of insurance namely automobile insurance, healthcare insurance, housing insurance, corporate insurance and so on.

An example of a current insurance fraud detecting techniques is a fuzzy support vector machine (SVM) model by Han Tao et al., [32] which based on traditional SVM with dual membership for the identification of insurance fraud. The empirical results show an impressive 90.73% precision on a recall of 91.31% when using dual membership fuzzy support vector machine model for fraud identification, which is higher compared to other models tested. Although, traditionally, insurance film uses statistical models to identify fraudulent claims. However, big data analytics addresses the big data challenges and plays a crucial role in fraud detection for insurance companies. Thus, big data applications are providing a faster, easier fruad solution for insurers.

11. Computer-based network financial fraud

When the individual is the target of cybercrime, the computer considered as the tool rather than the target. The example of both "computer as target" and "computer as a tool" in financial crimes include the use of the internet, mobile phone, malware, local area network, wide area network, wireless network, digital material and computer’s peripheral platform to steal money. The stealing may take place with the use of credit card, information or data leak, cyber-attack, viruses attack, hacking data, phishing, exploiting electronic financial systems and so on. However, the target of computer-based financial crime may be an individual, organizations, businesses and government agencies as it affects everyone. Thus, this method of crime can be referred to economic espionage, web defacement, sabotage of data, fraud and unauthorized access to a computer network.

Cybercrime is criminal activities carried out using computers or the Internet. Computer fraudsters prey on identities and banking passwords that are easy to steal. The risk of cyber-crime grows into sharp focus after tens of thousands of people across about one hundred (100) countries were affected by a massive cyber-attack in mid-May 2017 [35]. According to a new research banking and investment management professionals think that the most significant threat of financial crime now stems from cyberspace [35,36].

12. Other related types of financial fraud

Other related types of financial fraud are as follow;

(i) Occupational fraud: the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organizations.

(ii) Job bidding/tender fraud: involve advertising for procurement materials or contracts that are capital intensive but fake and require the companies to pay before bidding. The money collected in millions and no job issue out to any business.

(iii) Credit card skimming: this involves stealing data off a credit card during a legitimate transaction. Fraudsters swipe the card in an electronic device known as a "skimming device" or "wedge" which records all information contained on the magnetic strip.

(iv) International lottery fraud: the perpetrator may send a cheque which a target person is instructed to deposit in the account and to send the money back to the lottery. The perpetrator will create a "sense of urgency," compelling the targeted person to send the money before the cheque, which is counterfeit, is returned.

(v) Wills and legacies fraud: a letter or email sent to a target person claiming that someone has died and had mentioned a target person's name in the will. Usually, the scammer will claim to be the deceased's legal advisor and may request an advance fee.

(vi) Inheritance scams: involves a person receiving an email from an 'estate locator' or 'research specialist' purporting an unclaimed inheritance or refund. A targeted person lured into sending a fee to receive information about how to obtain the purported asset.

(vii) Fake prizes: a perpetrator claims that a person has won a non-existent prize. Then, the person asked to send a cheque to pay the taxes or provide credit card details or account number to pay for shipping and handling charges to send the non-existent prize.

(viii) Fund transfer scams: a target person may be asked through an advert or email to receive a payment into your bank account, and to send it abroad in return for receiving a commission.

(ix) Advance fee scams: these scams perpetrated through a letter, email or phone call offering a target person a large sum of money if the person can help someone transfer millions of rupees or other currency out of his country. The target person is asked to send details of person's bank account and an administration fee.

(x) Market manipulation: it referred to as "Pump and Dumps," it involves generating artificial buying pressure for a victim security, generally a low-trading volume issuer in the over-the-counter securities market that controlled through the fraud perpetrators. The artificially increased trading volume has the effect of insincerely raising the price of the targeted security (i.e., the Pump), which can be quickly sold off into the inflated market for the security through the fraud perpetrators (i.e., the Dump).

(xi) Broker embezzlement: these schemes involve unlawful and unauthorized actions by stockbrokers to steal directly from their clients. Such schemes may be facilitated by the doctoring of account statements, forging of client documents, illegal trading and funds transfer activities or other conduct in breach of the broker's fiduciary responsibilities to the victim client.

(xii) Late-day trading: schemes involve the unauthorized purchase and sale of securities after regular market hours for financial gains.

(xiii) Prime bank investment fraud: in these schemes, perpetrators claim to have access to a secret trading program endorsed by large financial institutions.

(xiv) Promissory notes: these are short-term debt instruments issued by partially known or non-existent companies.

(xv) Affinity frauds: perpetrators of affinity frauds, take advantage of the tendency of people to trust others with whom they share similarities such as religion or ethnic identity to gain their trust and money.

(xvi) Tax evasion fraud: is the illegal dodging of taxes by individuals, corporations, and trusts. Tax evasion often entails taxpayers deliberately false reporting the actual to the tax authorities to reduce tax liability and add dishonest tax reporting such as declaring less income, gains or profits than the amounts earned or overstating deductions.

(xvii) Financial statement fraud: is deliberate false reporting, misstatement or omission of financial statement data for misleading the evaluator or reader and creating a false impression of an organization's financial strength.

(xviii) Investment fraud: These schemes, sometimes referred to as High Yield Investment Fraud, involve the illegal sale or purported sale of financial instruments.

(xix) **Ponzi schemes fraud:** is a deceptive kind of investment fraud that includes the payment of falsely returns to existing participants or investors from money contributed by newcomer investors. However, in a short-term, the business is put into abrupt by perpetrators for financial gain.

(xx) **Pyramid schemes:** The Pyramid schemes are like Ponzi schemes because in a Ponzi the money collected from new investor or participants is paid to earlier participants or investors. However, in a pyramid scheme participant receive commissions for recruiting new participants into the scheme.

(xxi) **Borrowing fraud:** This involves borrowing of money from an individual or an organization and refuses to pay back or with the deliberate intention of not paying back.

(xxii) **Purchasing/service fraud:** this is a fraud of deception whereby an individual is given money to buy an item or a product or to render service. However, that individual neither buys the product or render the service and at the same time did not return the money to the person or client or an organization that gives out the money.

(xxiii) **Borrowing with interest fraud:** this fraud involves borrowing of money from an individual or an organization based on interest to execute a contract or deception to use the money for contract job but refuse to pay back the interest as agreed or refuse to pay back the money and the interest.

(xxiv) **Co-operative investment fraud:** These schemes, sometimes referred to as “problem-solving investment opportunity”, it involves an illegal contribution with rotational packing system among the contributors. However, in a short-term, the business is put into abrupt by perpetrators for financial gain.

13. The growth of big data techniques for financial crime detection in the literature

In this survey paper, the materials are obtained from the highly-reputed publishers such as Elsevier, Springer, and IEEE as well, from other journals which were searched using Google Scholar. This state of the art survey has shown that financial fraud has attracted the most attention from researchers in recent time. The evaluation of the selected articles reveals that the opportunity clearly exists to strengthen empirical research based on in-depth study case based quantitative approach and qualitative, as most of the articles analyzed followed an analytical method.

The result indicates that anomaly, data-mining, neural networks, clustering and outlier detection techniques are most commonly used techniques for financial fraud detection. Though all the methods could be applicable to one problem and the other, meanwhile some are mostly adaptable to any challenges in the financial domain than another. However, this outcome does not necessarily mean that they are the most effective or efficient in practice. As it seems, there is an algorithm, model, scheme or framework for every problem in the society. The result of the reviewed framework for fraud classification is shown in Table 1.

Table 1: Summary of the result of the reviewed framework on BDA for financial crime detecting classification.

Techniques//Methods/Approaches	Financial Crimes						Total
	Bank fraud	Security and exchange commodities fraud	Insurance fraud	Money laundering fraud	Computer network fraud	Other related fraud	
Anomaly	3	2	1	1	2	1	10
Neural network	2	1	1	2	1	1	8
Hybrids features	2	1	1	1	2	1	7
Models and algorithms	1	1	1	1	1	1	6
User’s behaviours and transaction histories	1	1	1	1	1	1	6
Computing paradigms	1	1	1	1	1	1	6
Intelligent control model	1	1	1	1	1	1	6
Quantitative framework	1	1	1	1	1	1	6
Statistical methods	1	1	1	1	1	1	6
machine learning models	1	1	1	1	1	1	6
XBRL technology	1	1	1	1	1	1	6
OLAP query languages	1	1	1	1	1	1	6
Classification	1	1	1	1	1	1	6
Cryptography algorithms	1	1	1	1	1	1	6
Change point detection,	1	1	1	1	1	1	6
Non-parametric	1	1	1	1	1	1	6
Hoeffding’s inequality,	1	1	1	1	1	1	6
Bernstein’s inequality	1	1	1	1	1	1	6

M-I divergence estimation	1	1	1	1	1	1	6
Clustering	2	1	1	1	2	1	8
Outlier detection	2	1	1	1	1	1	7
Granger causality test	1	1	1	1	1	1	6
VAR	1	1	1	1	1	1	6
Levene’s test homoscedasticity	1	1	1	1	1	1	6
SQL extensions,	1	1	1	1	1	1	6
Graphs paths	1	1	1	1	1	1	6
Alerts hat contact the cardholder channel voice	1	1	1	1	1	1	6
Alerts hat contact the cardholder channel SMS	1	1	1	1	1	1	6
Alerts hat contact the cardholder channel email	1	1	1	1	1	1	6
Stage-wise filtering system	1	1	1	1	1	1	6
Genetic programming	1	1	1	1	1	1	6
Success rate	1	1	1	1	1	1	6
Performance metrics	1	1	1	1	1	1	6
Agent-based simulation	1	1	1	1	1	1	6
Mathematical statistics	1	1	1	1	1	1	6
Modelling	1	1	1	1	1	1	6
Correlation between financial crime and attributes transaction	1	1	1	1	1	1	6
Transactional and sentimental analysis	1	1	1	1	1	1	6
Intelligent risk models	1	1	1	1	1	1	6
Statistical data	1	1	1	1	1	1	6
Data mining	3	1	1	2	2	1	9
Data visualization,	1	1	1	1	1	1	6
Filtering tools.	1	1	1	1	1	1	6
Text mining techniques	1	1	1	1	1	1	6
Dirichlet allocation modeling	1	1	1	1	1	1	6
An activity-based intelligence (ABI) model	1	1	1	1	1	1	6
Density-based clustering algorithms	1	1	1	1	1	1	6
Regression analysis	1	1	1	1	1	1	6
Hybrid technologies and models analysis	2	1	1	2	1	1	8
Big data analytics with Hadoop framework analysis	1	1	1	1	1	1	6

14. Conclusion

The results of the study indicate that the selected articles were recent because they were mainly published between 2010 and 2017 and focuses on technological issues regarding big data techniques for detecting financial fraud. Diverse methods were proposed for data collection, storage, transport, processing and analysis in the selected studies. The effective of enhanced authentication methodology and system with deviceless OTPs [37] could provide users and financial customers with a simple, more secure method to access data leveraging their current passwords.

Although, Big data tools make it simple, easy and faster for data analysts [38-41] to enable not only the discovery of a new method for financial crime typologies but also cross-product, cross-channel and cross-customer crime. However, Big data is not necessarily a panacea; there has to be information and communication technology (ICT) involvement in making the data sources available and, of course, big data tools do not exclude the need for data cleansing. But it gives new opportunities to an organization looking to tackle financial crime via new means.

Lastly, an evaluation was made of the detection techniques based on data analysis factors such as processing speed, latency, volume, performance, fault tolerance, scalability, and accuracy. Then we propose that emerging techniques such as neural-networks, data-mining, anomaly, clustering, hybrid-technologies, neuro-fuzzy, *rough* clustering, rough C-means (RCM), and machine learning. Others including Genetic algorithms, k-*Means* clustering, fuzzy support vector machine (SVM) model and quantum computing [42-48] could perform better than other methods currently in practice.

15. Areas of further research

It would be a good idea for further research to be carried out on;

(i) The areas of big data infrastructures like hardware equipment, software licensing, and maintenance (because of the high level of technical skill) which are still very expensive.

(ii) Human analysis of big data techniques for financial crime detection because of the rigorous of sorting out information.

(iii) Integration of ICT tools like a robot, CCTV camera, social network analysis, data mining techniques into the training of financial institution employees and security personnel.

(iv) The area of emerging trends in big data techniques such as; statistical machine learning, spatial mining, sequential and temporal mining, mining high-speed data streams and sensor data, process mining, privacy-preserving data mining. Others include network mining, web mining, column-based DBMS, parallel DBMS, in-memory DBMS, cloud computing, Hadoop, MapReduce and so on and so forth could provide a significant result.

(v) Empirically illustration and validation theories of financial crime.

(vi) The performance of detection methods with respect to cost implication.

(vii) Research is requiring in the areas of big data infrastructures such as hardware equipment, software licensing, and maintenance which are still very costly.

(viii)Similarly, more research is needed into the human analysis of big data techniques for financial crime detection because of the rigorous of sorting out information.

Funding

This study was partially supported by two organizations. Center for Cyber Safety and Education, United States Internal Revenue segregated fund of (ISC)², Inc. Code. EIN: 45-2405127 through the (ISC)2 graduate cybersecurity scholarship award, 311 Park Place Blvd. Suite 610 Clearwater, FL 33759 USA. Also, by Fundamental Research Grant Scheme (FRGS) for ‘‘content-based analysis framework for better email forensic and cyber investigation” [203/PKOMP/6711426], SFRG Lab, School of Computer Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia.

References

[1] Tiwari, A., Knowles, J., Avineri, E., Dahal, K., & Roy, R. (Eds.). (2010). Applications of soft computing: Recent trends (Vol. 36). Springer Science & Business Media.

[2] Zicari, R. V. (2014). Big data: Challenges and opportunities. Big data computing, 564.

[3] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286.

[4] Wang, Y. and Wiebe, V. J., (2014). Big Data Analytics on the characteristic equilibrium of collective opinions in social networks, International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 8 (3) (2014), pp. 29-44.

[5] Akerkar R. Big data computing CRC Press, Taylor & Francis Group, Florida, USA (2014)

[6] Gottschalk, P. (2010). Theories of financial crime. Journal of Financial Crime, 17(2), 210-222.

[7] West, J., Bhattacharya, M., & Islam, R. (2014, September). Intelligent financial fraud detection practices: an investigation. In International Conference on Security and Privacy in Communication Systems (pp. 186-203). Springer, Cham.

[8] Conseillers en gestion et informatique (CGI) report 2017, Current Challenges in Fighting Financial Crime. https://www.cgi.com/en

[9] Reuters report, “Equifax data breach to cost insurers $125 million: Property Claim Services. 2 October 2017. www.equifaxsecurity2017.com

[10] B.B. Sagar, P. Singh, and S. Mallika, “Online transaction fraud detection techniques: A review of data mining approaches,” (INDIACom), 2016 3rd International Conference on, pp. 3756-3761. IEEE. 2016.

[11] M.L. Bhasin, “Corporate accounting fraud: A case study of Satyam Computers Limited,” International Journal of Contemporary Business Studies, vol. 3, no.10, October 2012.

[12] K. Pal, and T, Alpine, “How to combat financial fraud by using big data,” 2014. https://www.kdnuggets.com/2016/03/combat-financial-fraud-using-big-data.html

[13] I.H. Witten, E. Frank, M.A. Hall, and C.J. Pal, “Data Mining: Practical machine learning tools and techniques,” Morgan Kaufmann, 2016.

[14] S. Zamir, “The 5 most common types of financial crimes,” law office of Shahin Zamir http://www.criminallawyersinhouston.com/theftfraud/the-5-most-common-types-of-financial-crimes/

[15] Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. Information sciences, 8(3), 199-249.

[16] Verschae, R., del Solar, J. R., Koppen, M., & Garcia, R. V. (2005, November). Improvement of a face detection system by evolutionary multi-objective optimization. In Hybrid Intelligent Systems, 2005. HIS'05. Fifth International Conference on (pp. 6-pp). IEEE.

[17] Kouwenhoven, E. N., Van Heeringen, S. J., Tena, J. J., Oti, M., Dutilh, B. E., Alonso, M. E., & Bolat, E. (2010). Genome-wide profiling of p63 DNA–binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. PLoS genetics, 6(8), e1001065.

[18] A. Sudjianto, S. Nair, M. Yuan, A. Zhang, D. Kern, “Statistical Methods for Fighting Financial Crimes,” Technometrics, vol. 52 no.1, pp. 5-9, 2010.

[19] H. Chen, R.H.L. Chiang, and V.C. Storey, “Business Intelligence and Analytics: From Big Data to Big Impact,”. MIS Quarterly, 2012.

[20] P. Lemov, “Corporate Tax Attacks in the States At the same time states are looking to beef up corporate tax collections, they are also cutting corporate taxes,”

[21] B.B. Kelly, “Investing in a centralized cybersecurity infrastructure: Why hacktivism can and should influence cybersecurity reform,” BUL Rev., 2012

[22] S. Jung-a, “South Korea’s economy slows amid a cluster of corporate woes,” in Seoul, Financial Time, October 25, 2016.

[23] J. West, M. Bhattacharya, R. Islam “Intelligent financial fraud detection practices: An investigation,” ” International Conference on Security and Privacy 2014.

[24] D.A. Montague, "Essentials of online payment security and fraud prevention," John Wiley & Sons, vol. 54, 2010.

[25] N. Mansor, "Concomitant debacle of fraud incidences in the Nigeria public sector: understanding the power of fraud triangle theory," International Journal of Academic Research in Business and Social Sciences, 5(9), 241-255, 2015.

[26] *Bernama*, "60% bank fraud committed by employees," December 7, 2013. FMT News http://www.freemalaysiatoday.com/category/nation/2013/12/07/60-bank-fraud-committed-by-employees/

[27] Reuters, "Massive bank fraud trial starts in China," The Thomson Reuters Trust Principles Report 2007.

[28] D.Weinland, Chinese bank hit by $4.9billion loan fraud," Daily Financial Times News Report, 2016.

[29] J.R. Koren, "Korean investors sue Wilshire Bank over fraud that cost them millions," Los Angeles Times http://www.latimes.com/business/la-fi-Wilshire-eb5-fraud-20160725-snap-story.html

[30] J. Park, "South Korea reveals staggering $1 billion transfer fraud in Iranian money," Reuters Report.

[31] Jill Treanor, "UK fraud hits record £1.1bn as cybercrime soars," The Guardian News Report. 24th Tuesday, January 2017. https://www.theguardian.com/uk-news/2017/jan/24/uk-fraud-record-cybercrime-kpmg

[32] N. Raymond, and J. Stempel, "Deutsche Bank to pay $95 million to end U.S. tax fraud case," New York Reuters Report January 4, 2017.

[33] M. Levi, and P. Reuter, "Money laundering,"Crime and Justice, Crime and Justice, The University of Chicago Press, vol. 34, no. 1 pp. 289-375, 2006.

[34] The US. SEC, "SEC Announces Financial Fraud Cases," for immediate release, 74, Washington D.C., April 19, 2016. https://www.sec.gov/news/pressrelease/2016-74.html

[35] Wright, Lance. People, Risk, and Security: How to prevent your greatest asset from becoming your greatest liability. Springer, 2017.

[36] Johnson, K. N. (2017). 3. Innovating to new heists: regulating cyber threats in the financial services industry. The Most Important Concepts in Finance, 28.

[37] Omolara, A. E., Jantan, A., Abiodun, O. I., & Arshad, H. (2018). An enhanced practical difficulty of one-time pad algorithm resolving the key management and distribution problem. In proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1).

[38] Atyeh, A. J., Jaradat, M. I. R. M., & Arabeyyat, O. S. (2017). Big Data Analytics Evaluation, Selection and Adoption: A Developing Country Perspective. International Journal of Computer Science and Network Security, 17(9), 159-171.

[39] Abiodun, O. I., Jantan, A., Omolara, A. E., & Mahinderjit, M. M. (2018). Big Data: an approach for detecting terrorist activities with people's profiling. In proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1).

[40] AlMadahkah, A. M. (2016). Big data in computer cyber security systems. International Journal of Computer Science and Network Security (IJCSNS), 16(4), 56.

[41] H. He, and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on knowledge and data engineering, 2009, 21(9), 1263- 1284.

[42] Anbuvizhi, R., & Balakumar, V. (2016). Credit/Debit Card Transaction Survey Using Map Reduce in HDFS and Implementing Syferlock to Prevent Fraudulent. International Journal of Computer Science and Network Security (IJCSNS), 16(11), 106.

[43] Saeed, M. A., & Ahmed, K. (2017). Future of Data Security with the Emergence of Quantum Paradigm. International Journal of Computer Science and Network Security, 17(9), 1-5.

[44] Makki, S., Haque, R., Taher, Y., Assaghir, Z., Ditzler, G., Hacid, M. S., & Zeineddine, H. (2017, September). Fraud Analysis Approaches in the Age of Big Data-A Review of State of the Art. In Foundations and Applications of Self* Systems (FAS* W), 2017 IEEE 2nd International Workshops on (pp. 243-250). IEEE.

[45] Vanitha, R. (2016). Intrusion Prevention Against Distributed Denial-of-Service (DDoS) on the cloud. International Journal of Computer Science and Network Security (IJCSNS), 16(4), 90.

[46] Ubukata, S., Notsu, A., & Honda, K. (2017). General Formulation of Rough C-Means Clustering. International Journal of Computer Science and Network Security, 17(9), 29-38.

[47] Melo-Acosta, G. E., Duitama-Muñoz, F., & Arias-Londoño, J. D. (2017, August). Fraud detection in big data using supervised and semi-supervised learning techniques. In Communications and Computing (COLCOM), 2017 IEEE Colombian Conference on (pp. 1-6). IEEE.

[48] Duman, E. A., & Sağıroğlu, Ṣ. (2017, October). Heath care fraud detection methods and new approaches. In Computer Science and Engineering (UBMK), 2017 International Conference on (pp. 839-844). IEEE.