Journal of Health Informatics in Africa

10th Health Informatics in Africa Conference (HELINA 2017)
Peer-reviewed and selected under the responsibility of the Scientific Programme Committee

A Schematic View of the Application of Big Data Analytics in Healthcare Crime Investigation

Terungwa Simon Yange [a*], Hettie Abimbola Soriyan[b], Oluoha, O[c].

[a]Department of Mathematics/Statistics/Computer Science, Federal University of Agriculture, Makurdi, Nigeria
[b]Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
[c]Department of Computer Science, University of Nigeria, Nsukka, Nigeria

Background and Purpose: One major challenge encountered during crime investigation via automated systems is the inability of conventional data analysis techniques to adequately handle the enormous data that are made available during the investigation. Existing crime investigation frameworks are built on orthodox data analysis techniques which cannot sufficiently manned the unprecedented size and variety of data available today, not to mention the significantly more anticipated data in the near future. This has affected the healthcare industry where data is predominantly multi-structured and is growing at a considerably faster rate.
Methods: To address this, a big data analytics model based on deep learning was designed in this research using enterprise application diagrams.
Results: This model is intended to be implemented using Apache Hadoop a big data implementation framework. When implemented, the model will create a platform that will handle a phenomenon that is affecting millions of people all over the world.
Conclusions: This is the first of its kind to use big data analytics techniques in healthcare crime investigation in Nigeria which provided security intelligence by shortening the time of correlating and deriving evidence from large volume of data for healthcare crime investigation purposes. Finally, this research also enabled the healthcare systems to systematically use big data analytics to identify inefficiencies and best practices that improve care delivery and reduce costs.

Keywords: Crime, Hadoop, Deep Learning, Investigation Data Analytics, Health Insurance

1 Introduction

Big data usually includes datasets whose sizes are beyond using conventional data analysis tools to manage. The analysis of big data commonly known as big data analytics is the process of collecting, organizing and analysing large, diverse dataset that involves different types such as structured and unstructured, and streaming and batch, with sizes from terabytes to zettabytes to discover patterns and other useful information [1]. Big data analytics can be applied in information security which involves the ability to gather massive amounts of digital information to analyse, visualize and draw insights that can make it possible to detect crime. This can transform security analytics by improving the maintenance, storage and analysis of security information. Big data analytics correlate the data drawn from multiple sources such as network traffic, log files, financial transactions, healthcare claims etc. into a coherent view so as to identify the anomalies and suspicious activities of the criminals. Big data is ideal for investigating information security issues; and detecting a crime is largely about uncovering data patterns that are not ordinary from log files. Applying big data techniques will ease such analysis to reveal anomalies that point to a data breach. With this powerful strength, big data analytics could lead to the discovery of big crime which invariably could culminate into 'big arrest'.

*Corresponding author address: Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi, Nigeria, Email: lordesty2k7@ymail.com, Tel: +2348064067803

J Health Inform Afr. 2017;4(1):32-41. DOI: 10.12856/JHIA-2017-v4-i1-161

Crime encompasses a wide range of illicit practices and illegal acts involving intentional deception or misrepresentation. Crime is any illegal act characterized by deceit, concealment, or violation of trust [2]. These acts are not dependent upon the threat of violence or physical force. Crimes are perpetrated by parties and organizations to obtain money, property, or services; to avoid payment or loss of services; or to secure personal or business advantages." In other words, crime is a harmful act or omission against the public which the society wishes to prevent and which, upon conviction, is punishable by fine, imprisonment, and/or death. No conduct constitutes a crime unless it is declared criminal in the laws of the country [3][4]. Some crimes (such as theft or criminal damage) may also be civil wrongs (torts) for which the victim(s) may claim damages in compensation. Crime and fraud are synonymous, therefore in this research, the two words will be used interchangeably.

Crime impacts organizations negatively in several areas including financial, operational, and psychological. While the financial loss owing to crime is significant, the full impact of crime on an organization can be overwhelming. The losses to reputation, goodwill, and customer relations can be also devastating. The society is strongly affected by crime, both due to the cost of crime, as well as the decline in the quality of life that citizens suffer as a consequence of crime. As crime can be perpetrated by any employee within an organization or by those from the outside, it is important to have an effective crime management programme in place to safeguard the organization's assets and reputation. Crime and society are closely linked-for better and for worse and is as old as humanity, and occurs in different degrees of severity. However, society can also play a role in reducing and deterring crime. Many agencies and programmes in crime management are based on societal and community efforts. The magnitude of criminal activities can be perceived in all spheres of life [2].

For instance, the healthcare sector is among the most information intensive industries. Its information, knowledge and data keep growing on a daily basis and the ability to extract useful information that will improve the quality of healthcare services rendered is very crucial. Crime in this sector involve the intentional deception or misrepresentation for gaining some shabby benefits in the form of health expenditures [4]. This can be anything like providing false and intentionally misleading statements to patients, submitting false bills or claims for services, falsifying medical records or reports, lying about credentials or qualifications, unnecessary medical treatment or drug prescription; which seriously drain the finances in the healthcare system. This severely deters the healthcare industry from providing quality and safe care to legitimate patients; and it has called for an effective crime management system so as to reduce this illegal behaviour with the intention of improving the quality and reducing the cost of healthcare services. Owing to the large number of cases reported, investigated and prosecuted, it has been identified as a "high-risk" area in many regions such as the UK, the US, Romania, Nigeria etc. [4].

Healthcare crime exist in many forms: dishonest providers, organized criminals, collusion with patients, and patients who misrepresent their eligibility for health insurance coverage. It can be categorized into: health insurance crime, drug crime and medical crime. Due to the confidentiality of the medical records, data for healthcare crime comes mostly from health insurance crime; and it occurs when a company or an individual defrauds an insurer or government healthcare programme. In this paper, a survey of the existing healthcare investigation approaches is carried out and a new approach is designed.

1.1 Related Works

Crime in the healthcare insurance involve three parties [4][5]: healthcare service provider (i.e., the physician, pharmacist, laboratory scientist, healthcare centre, pharmacy, laboratory, and even ambulance companies) which render healthcare services; healthcare service consumer or beneficiary or insurance subscriber (i.e., patient) which receive healthcare service from the provider; and the healthcare insurance carrier which receive regular premiums from subscribers and make the commitment to pay healthcare cost on their behalves. These parties exchange information amongst them in the course of care delivery. This is basically in the form of service requested by the subscriber (patient visit) to the provider, explanation of benefits which contain the detail services rendered by the provider to the subscriber, claim/bill which is sent to carrier for the services rendered to the subscriber by the provider, and the payment to the provider based on the claim submitted to the carrier. As the number of beneficiaries (patients) of this scheme increases, high volume of data is generated by both the providers and the carriers; and consequently, some fraudulent activities (such as billing services that were never rendered, performing medically unnecessary services, misrepresenting non-covered treatments as medically necessary covered treatments, and misrepresenting applications for obtaining lower premium rate) are carried by these actors (beneficiary, provider and insurer) which give rise to the need to investigate such acts in an attempt to identify perpetrators, and this requires a proper analytics tool for the purpose [4][6].

Recent development of new technologies eased production, collection and storage of high dimensional and complex data. Healthcare has been no exception. Modern medicine generates a great deal of data which is stored in medical databases. Medical databases are increasing in size in three ways [6]: the number of records in the database, the number of fields or attributes associated with a record, and the complexity of the data itself. Extracting pertinent information from such complex databases for inferring potential fraudulent activities has become increasingly important for fraud detection. [4] gives an account of the amount of information involved in the reimbursement process for healthcare insurance scheme, which supports the cost of prescription medications to seniors and the disabled in the US. In such a complex process, involving many actors, the possibility of fraud cannot be overlooked. At the same time, quality of medical records should be ensured to avoid, for instance, fraudulent claims. Also, there is yet another type of healthcare crime called "conspiracy fraud" which involves more than one party [5]. An important characteristic of conspiracy fraud is the need to deal with dyadic data connecting the involved parties. The important feature of dyadic data is that it can be organized into a matrix where rows and columns represent a symmetric relationship. In healthcare fraud detection, the typical relationship of interest is the one between a provider and a beneficiary.

With this enormous amount of data that is generated in the healthcare industry, traditional methods of detecting healthcare crime are time-consuming and inefficient due to the complicated nature of medical processes and the complexity of the data have made crime to have a favourable niche in the healthcare systems as most crimes go undetected. Conventional analysis methods are not suitable due to their inner limitations to manage the volume, velocity, variety, veracity, value and complexity of the data in the healthcare [5][7][8][9][10][11].

Among these three types of fraud, the one committed by health service provider's accounts for the greatest proportion of the total healthcare fraud. Although a vast majority of service providers are honest and ethical, but a few dishonest ones may have various possible ways to commit fraud on a very broad scale, thus posing great damage to the health care system. Some service providers' fraud, such as that involving medical transportation, surgeries, invasive testing, and certain drug therapies, even places patients at a high physical risk.

1.2 Healthcare Crime Investigation Approaches

[11] proposed the use of cluster analysis for geographical analysis of potential fraud. The emphasis of this work was on types of fraud committed by a single party. But as pointed out by [5] some frauds in the health insurance involves more than one party: conspiracy or conspiratorial frauds. A typical conspiratorial fraud scenario is that patients collude with physicians, fabricating medical service and transition records to deceive the insurance company to whom they subscribed. This can be very rewarding owing to its complexity, increasing popularity, and severe consequences. An important characteristic of conspiratorial fraud is the need to deal with dyadic data connecting the involved parties. The important feature of dyadic data is that it can be organized into a matrix where rows and columns represent a symmetric relationship. In healthcare fraud detection, the typical relationship of interest is the one between a provider and a beneficiary. As also noted by [5], detection of conspiracy fraud has not gained much attention in the health care fraud literature. In what follows, [12] considered the use of co-clustering methods for detection of conspiracy fraud. In so doing, the proposed models were able to describe and capture the dyadic dynamics that connects providers and beneficiaries. Co-clustering allows the grouping of providers and beneficiaries simultaneously, that is, the clustering is interdependent.

[9] developed a data mining technique for fraud detection in health insurance scheme using knee-point k-means algorithm. They considered NHIS as the case study for the work. The work, focuses on the application of some computer-based techniques that could help to properly target investment in the healthcare sector and also drastically reduce fraud in health insurance by healthcare providers. To this effect, they applied the knee-point k-means clustering method, which was capable of detecting fraudulent claims by health service providers. Cluster-based outliers were examined. Health providers' claims submitted to a HMO were grouped into clusters. Claims with similar characteristics were grouped together. The claims were grouped into two clusters: fraudulent and non-fraudulent. The results from the

data collected from a particular HMO in Lagos, Nigeria show that the total number of claims identified as possible anomalies from cluster-based outliers was seven (7) in Nigeria health insurance using probability of 0.6 as the cut-off point. This research did not classify the fraud detected, whether it is provider, consumer or insurer frauds; it uses only the unsupervised technique (K-Means algorithm) for clustering; and the data was collected from only one HMO which cannot yield a perfect result.

In a survey on hybrid approaches for fraud detection in health insurance by [7], the act committed with the intent to obtain a fraudulent outcome from an insurance process was carefully examined. According to them, when a claimant attempts to obtain some benefits or advantages to which they are not entitled then that attempt is considered as insurance fraud and it has become a major concern for health insurance companies. They proposed a hybrid framework that applied some data mining techniques to detect frauds. This framework considered the analysis of the characteristics of healthcare insurance data, some preliminary knowledge of healthcare system and the fraudulent behaviours. The framework harnessed the advantages of both the supervised and unsupervised learning techniques to detect fraudulent claims. This framework did not consider the high volume, velocity, variety, veracity etc. of data and it was not implemented.

In the same vein, [13], investigated the benefits of big data technology and the main methods of analysis that can be applied to the cases of fraud detection in public health insurance system in Romania. They outlined the benefits of using big data technology in combating crime in the healthcare industry.

Consequently, methods for identifying and preventing fraud must always be adjusted and ready to rediscover the fraudulent actions [3][4]. To add to the lapses in legislation of the country, each country has unique economic, political, social, and institutional opportunities for and barriers which makes fraud examination different amongst countries. A crucial and peculiar issue in the Nigerian National Health Insurance Scheme is the high level of corruption in the sector, lack of accountability and clear sense of irresponsibility [14].

[15] proposed a model using big data in investigating real time crime in the health insurance in the cloud. This approach utilizes fraud management solution to detect potential frauds in the cloud. The solution was based on a high volume of historical data, predictive statistical models and social media analytics. It renders its services through client components like apps and web-services. Just like the [7][13][15] did not implement any working system for their research. Also, as opined by [13], healthcare crimes are country specific and Nigeria has not adopted the cloud services and there is no healthcare laws relating data on the cloud, and therefore, this model cannot be used to investigate crime in our healthcare system.

In [10], an approach of data mining implementation in medical fraud detection was carried out. They considered the increasing amount of data stored in files, databases, and other repositories which requires the development of a powerful means to analyse and extract interesting knowledge from them. The healthcare fraud detection requires compilation of potentially huge data, involving complex computation and sorting operations. Once such frauds have been detected and classified, data cleaning is applied to it which helps to remove the noise and inconsistencies in the data thereby enhancing its quality. This technique can be used to detect the sale of potentially dangerous medicine by pharmacists thereby preventing such medical fraud.

[4] in their bid to address these issues provides an approach to detect and predict potential frauds by applying big data, Hadoop environment and analytics methods which led to rapid detection of claim anomalies. The solution was based on a high volume of historical data from various insurance company data and hospital data of a specific geographical area. With the voluminous, diverse, and varying nature of the data used, the distributed and parallel computing tools collectively termed big data were employed. The work demonstrated the effectiveness and efficiency of the open-source predictive modelling framework used to describe the results from predictive model. The research was able to detect erroneous or suspicious records in submitted healthcare datasets and proved how the hospital and other healthcare data are helpful for the detecting healthcare insurance fraud. The research also used the decision tree algorithm.

In [16], a fraud detection approach in the health insurance using data mining techniques was developed. This approach used SVM (Support Vector Machine) and Evolving Clustering Method (ECM) in health insurance field for fraud detection. In the research, SVM algorithm was used for classification and ECM algorithm was used for clustering. The SVM was used to train the system to determine decision boundary between legitimate and fraudulent claims classes while the ECM identifies new data point that comes in, it clusters them by modifying the position and size of the cluster (i.e., used to cluster dynamic data hence it find out newly incoming fraudulent claims).

[17] developed a model for detecting healthcare fraud and abuse using the supervised and unsupervised data mining techniques. According to them, the supervised methods applied to healthcare fraud and abuse detection are decision tree, neural networks, genetic algorithms and Support Vector Machine (SVM); while the unsupervised methods that have been applied to health care fraud and abuse are clustering, outlier detection and association rules. They concluded that outlier detection is an unsupervised method and routine online processing task as supervised learning method.

In our research, we considered the further classification of fraud so as to report the actual fraud. We also considered deep learning which combined both supervised and unsupervised techniques, since hybrid methods are proven to yield better results [7][16][17][18]. We collected data from different stakeholders in the healthcare insurance relating to fraud. This research also used big data analytics techniques to carry out the investigation since the anticipated data is very large in size.

1.3 Nigerian National Health Insurance Scheme

A beneficiary got enrolled in the National Health Insurance Scheme (NHIS) programme after an eligibility screening is performed by the NHIS. The person's circumstances and income is verified and if the criteria are met, the person can enrol in the NHIS programme. When the beneficiary is ill, he/she will go to the hospital where he/she will get the necessary treatment. The provider of the services provided to the beneficiary would submit the bill to the Health Maintenance Organization (HMO) in the form of claims. It is generally assumed that all of the providers have an agreement with NHIS and that they participate in the programme. The providers send the claim to the HMO which is reviewed and processed for the payment of the services of participating providers to the beneficiaries.

An Explanation of Benefits (EOB) is sent to the beneficiary; an EOB contains an overview of the provided services. This is an automatically generated detailed overview of the provided services and the corresponding codes and amounts. The claim submission and processing is done manually which makes most frauds go undetected. Verification of the legitimacy of claims become more tasking creating the assumption that all claim submitted are true and genuine. If a claim is rejected, there is no follow up to investigate why the provider submitted this claim. The said provider receives the information with the explanation why the claim was rejected. The lesson how not to submit the claim can be mastered and in the next attempt taken into account. Because the perpetrators are not investigated, they get wiser by the note that is sent to them and learn about the billing rules.

To investigate this process, provider should submit additional documents to prove that the services were actually provided because if he/she provided the service it is not a lot of effort to submit additional documentation. However, a fraudulent provider has a good reason not to reply; the fraud attempt failed and there are no consequences when he/she does not reply since there is no investigation. However, if an honest provider made a mistake with the submitted claim he/she would send the additional documents to get his/her provided services reimbursed. Further investigation of rejected claims will reduce most of the frauds in this scheme and the NHIS can benefit of this information source.

2 Methodology

This paper employed the hierarchical structure of the deep learning (deep belief network) architecture will be employed to learn high level representations and complicated structure automatically from complex health insurance data collected from the various health facilities, and stored in the Hadoop Distributed File System; and also provide algorithms that will be implemented by the MapReduce Programming platform for the distributed processing of the data. It provides a framework that allows distributed processing of complex datasets using simple programming models. The multiple layers of this architecture continuously abstract features of the data undergoing processing from one layer to another, making the search for fraudulent claims simpler. The pictorial view of this model was designed by enterprise application diagram. It uses the Nigerian Health Insurance Scheme (NHIS) as a test bed.

DOI: 10.12856/JHIA-2017-v4-i1-161

3 Results

This paper studies how the big data framework can be leveraged to extract, preprocess and analyse data from the NHIS with the aim of identifying fraud. Apache Hadoop is the Big Data framework considered. The Hadoop Distributed File System (HDFS) implementation of the Hadoop is used as an alternative to store data and the MapReduce to process extremely large data sets on commodity hardware. In addition, we will use Hive as an open-source data warehousing solution which is built on top of Hadoop. The design of the system is shown in figures 1 and 2. Hive supports queries expressed in a SQL like declarative language-HiveQL. RHadoop is a bridge between R, a language and environment to statistically explore data sets, and Hadoop, a framework that allows for the distributed processing of large data sets across clusters of computers. In this research, we will focus on using the ability of big data analytics to manage millions of events efficiently to develop a hybrid model for investigating health insurance fraud. The proposed model will be able to identify the erroneous or suspicious insurance claims during investigation. The conceptual view of the model is shown in figure 1 and figure 2. The model took into cognizance the high volume of data from this sector.

4 Discussion

The model designed in this paper when implemented will harness the big data's capability of crime investigation which will be of great value especially in the healthcare insurance scheme. This will be able to investigate breaches in security, determine compliance with established policies and operational procedures, and enable the reconstruction of sequences of events affecting the healthcare insurance domain to enable auditors do their work efficiently. It considered the volume and complexity of this data which made it impossible for humans and other traditional means to be sufficient enough to identify the crime perpetrated by the hoodlums in the healthcare insurance industry. The deep belief network algorithm of deep learning was used to "learn" normal activities so as to fish out any unwholesome activity in the healthcare insurance. The model created room to capture and process the data, help to visualize its flow and apply automatic learning techniques capable of discovering patterns and detecting anomalies from such patterns for proper investigation of the activities of fraudsters. With this, common repetitive errors that are "hidden" inside huge repositories of data which would go undetected in the absence of big data technologies because the orthodox techniques not being capable to correlate the huge quantities of data available in the medical sector will easily be identified and corrected. The deep learning architecture combined two machine learning theories: unsupervised and supervised theories, one is used in the pretraining while the other is used in fine-tuning the network. In the pretraining of a deep belief network, the unsupervised learning theory is used. This is aimed at finding clusters of similar inputs in data without being explicitly told that these data points belong to a different class. With the aid of this theory, unlabelled NHIS data will be used in to initialize the network in the pretraining phase. The supervised theory is aimed classifying inputs data with the aid of the target output. This theory implements the Back Propagation (BP) Algorithm which expressed the logic behind it. The idea behind BP algorithm is quite simple, output of the network is evaluated against desired output. If results are not satisfactory, connection (weights) between layers are modified and the process is repeated again and again until the error is small enough to be ignored. This theory aids the fine-tuning of all the weights and biases in the network pretrained by the unsupervised theory.

As opined by [4][7][17] and other researchers, this combined hybrid approach in the deep learning makes it easier to detect erroneous or suspicious records in submitted healthcare datasets and proved how the hospital and other healthcare data are helpful for the detecting healthcare insurance fraud. Components of the system are discussed below.

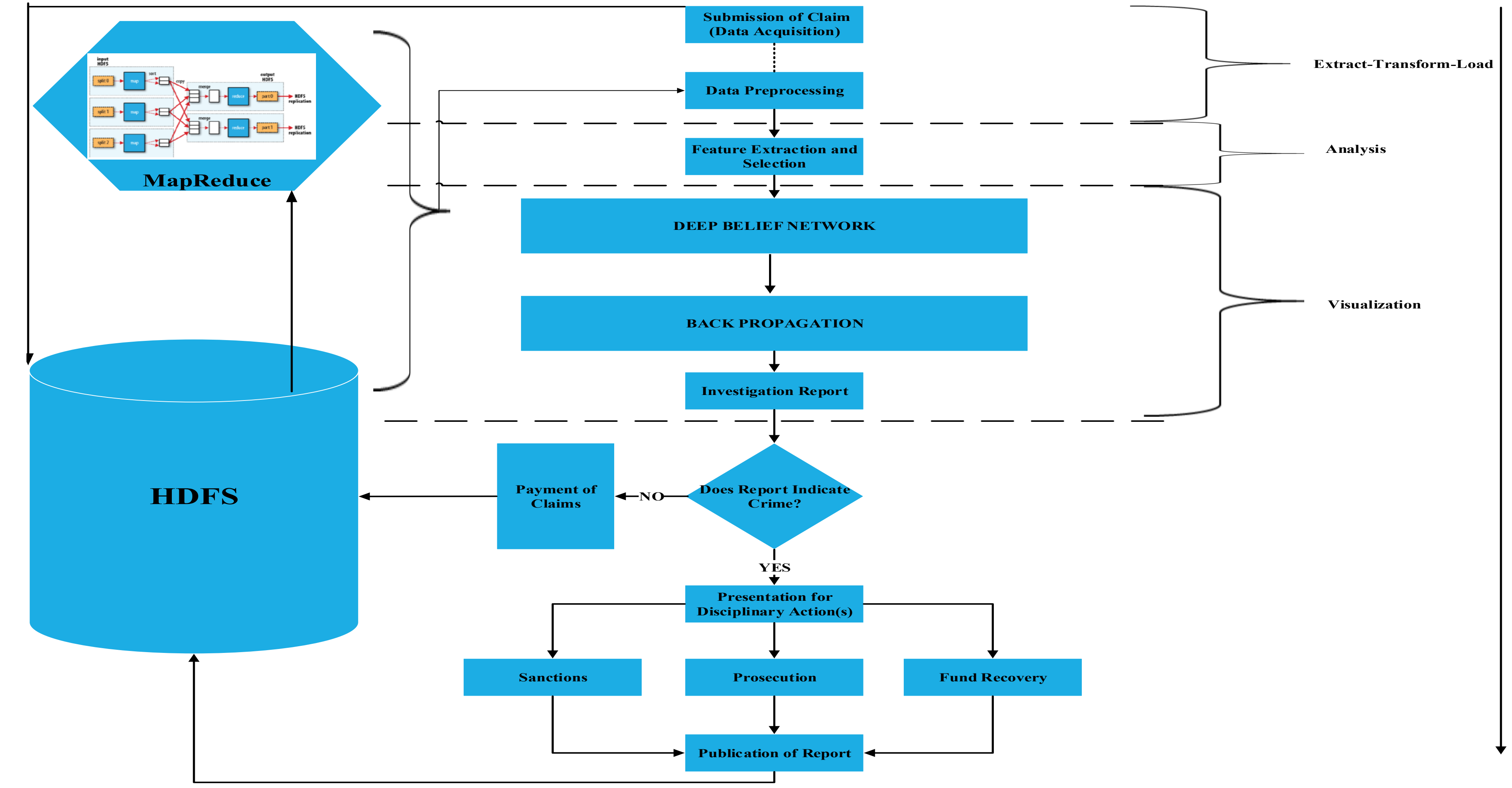

Figure 1. Schematic View of the Model

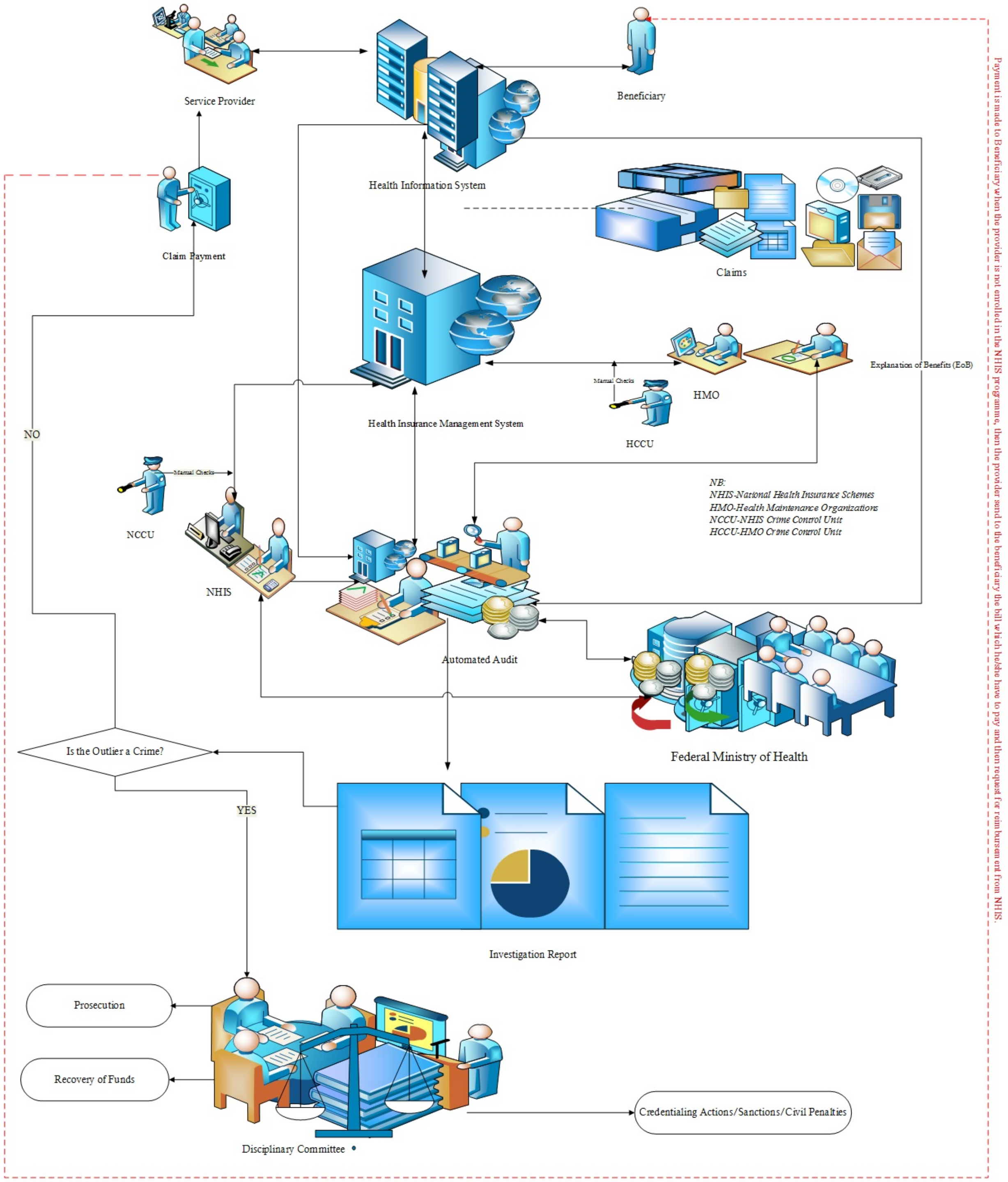
Service Provider
Beneficiary
Health Information System
Claim Payment
Claims
Health Insurance Management System
Manual Checks
HMO
Explanation of Benefits (EoB)
HCCU
NO
Manual Checks
NCCU
NHIS
NB:
NHIS-National Health Insurance Schemes
HMO-Health Maintenance Organizations
NCCU-NHIS Crime Control Unit
HCCU-HMO Crime Control Unit
Automated Audit
Federal Ministry of Health
Is the Outlier a Crime?
YES
Investigation Report
Prosecution
Recovery of Funds
Disciplinary Committee
Credentialing Actions/Sanctions/Civil Penalties
Payment is made to Beneficiary when the provider is not enrolled in the NHIS programme, then the provider send to the beneficiary the bill which he/she have to pay and then request for reimbursement from NHIS.

Figure 2. Model Flow

Hadoop Framework: This is the combination of the MapReduce and the Hadoop Distributed File System (HDFS). This is the backbone of this model. It enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that both store and process the data, and can scale without limits. It can handle all types of data from disparate systems: structured, unstructured, log files, pictures, audio files, communications records, email - regardless of its native format. Even when different types of data have been stored in unrelated systems, it is possible to store it all into Hadoop cluster with no prior need for a schema.

Data Acquisition and Preprocessing: Real world data that is collected from different sources is noisy and heterogeneous (different format) in nature. The heterogeneity in the healthcare data is responsible for the prevalence of missing values and inconsistencies which poses a great challenge leading to an inaccurate result if not addressed at the beginning. Raw data must be processed (this task is associated with segmentation, normalization and noise removal algorithms) into a form that is acceptable. This helps in constructing the homogeneous data set.

Analysis: This step aims to define new features out of the original attributes, to maximize the discrimination power of the machine learning method in separating fraudulent and legitimate cases. The feature extraction and selection procedures aim at finding the minimum number of discriminative features that are considered. The term feature selection refers to algorithms that select the best subset of the input feature set, whereas methods that create new features based on transformations or combinations of the original feature set are called feature extraction algorithms. Feature extraction takes in a pattern and produces features values. Feature extraction may provide a better discriminative ability than the best subset of given features, but these new features (a linear or a nonlinear combination of given features) may not have a clear physical meaning. Health insurance is a complex phenomenon governed by multiple variables because there is no universal factor that can be used to predict the fraud.

Visualization: Visualization present large volumes of data, provide interactivity to explore the data, make visual patterns easy to see and make multivariate analysis simple and easy to comprehend.

Other Components: The investigation report is produced after this stage. If the report does not indicate any fraudulent act, payment is made, but if it is otherwise, the report is presented for disciplinary action (sanction, prosecution in the court of law and fund recovery) from the disciplinary committee. This stage marks the end of the investigation and everything is recorded in the database. Depending on the success of an investigation, further action may result in the form of civil cases, criminal prosecutions or both. Such actions can result in monetary penalties (fines), recoveries of funds (belonging to the public trust or to private payers) and industry sanctions (such as revocation of privileges or prohibition against administering services to NHIS patients) against offenders.

5 Conclusion

This research designed a model that will automatically identify the general patterns of suspicious behaviour of criminals in the healthcare insurance claims. It employed advance data analysis techniques of big data which aid auditors in investigating breaches in security, determine compliance with established policies and operational procedures, and enable the reconstruction of sequences of events affecting the healthcare insurance domain. This brought humans and computers to collaborate and work closely together in crime investigation. With this, it will reduce the time it takes to uncover fraudulent activity, and also shrink the negative impact of significant losses owing to fraud. It has created a platform that will handle a phenomenon that is affecting millions of people all over the world. This is the first of its kind to use big data analytics techniques in healthcare crime investigation in Nigeria which provided security intelligence by shortening the time of correlating and deriving evidence from large volume of data for healthcare crime investigation purposes. Finally, this research also enabled the healthcare systems to systematically use big data analytics to identify inefficiencies and best practices that improve care delivery and reduce costs.

J Health Inform Afr. 2017;4(1):32-41. DOI: 10.12856/JHIA-2017-v4-i1-161

References

[1] Ularu EU, Puican FC, Apostu A, Velicanu M. Perspectives on Big Data and Big Data Analytics. DSJ. 2012' 3(4): 3-14.

[2] IACA (International Association of Crime Analysts). Definition and Types of Crime Analysis. White Paper released by Standards, Methods, & Technology (SMT) Committee. 2014.

[3] Yunusa U, Irinoye O, Suberu A, Garba AM, Timothy G, Dalhatu A, Ahmed S. Trends and Challenges of Public Health Care Financing System in Nigeria: The Way Forward. IOSR-JEF. 2014; 4(3): 28-34.

[4] Dutta A, Hongoro C. Scaling Up National Health Insurance in Nigeria: Learning from Case Studies of India, Colombia, and Thailand. Washington, DC: Futures Group, Health Policy Project. 2013.

[5] Dora P, Sekharan GH. Healthcare Insurance Fraud Detection Leveraging Big Data Analytics. IJSR. 2015; 4(4): 2073-2076.

[6] Li J, Huang K-Y, Jin J, Shi J. A survey on statistical methods for health care fraud detection. HCMS. 2008; 11: 275-287.

[7] Ekin T, Ieva F, Ruggeri F, Soyer R. Applications of bayesian methods in detection of healthcare frauds. CET. 2013; 33: 151-156.

[8] Bagul PD, Bojewar S, Sanghavi A. Survey on Hybrid Approach for Fraud Detection in Health Insurance. IJIRCCE. 2016; 4(4): 6918-6922.

[9] Bagde PR, Chaudhari MS. Analysis of Fraud Detection Mechanism in Health Insurance Using Statistical Data Mining Techniques, IJCSIT. 2016; 7(2): 925-927.

[10] Fashoto SG, Owolabi O, Sadiku J, Gbadeyan JA. Application of Data Mining Technique for Fraud Detection in Health Insurance Scheme Using Knee-Point K-Means Algorithm. AJBAS. 2013; 7(8): 140-144.

[11] Jacqulin MJ, Shrijina S. Implementation of Data Mining in Medical Fraud Detection. IJCA. 2013; 69(5): 1-4.

[12] Musal R. Two models to investigate Medicare fraud within unsupervised databases. ESA. 2010; 37(12): 8628-8633.

[13] Travaille P, Thornton D, Müller RM, Hillegersberg J. Electronic Fraud Detection in the U.S. Medicaid Healthcare Program: Lessons Learned from other Industries. Proceedings of the Seventeenth Americas Conference on Information Systems, Detroit, Michigan August 4th-7th 2011.

[14] Bologa A, Bologa R, Florea A. Big Data and Specific Analysis Methods for Insurance Fraud Detection. DSJ. 2010; 1(1): 30-39.

[15] Agba MO, Ushie EM, Osuchukwu NC. National Health Insurance Scheme (NHIS) and Employees' Access to Healthcare Services in Cross River State, Nigeria. GJHSS. 2010; 10(7): 9-16.

[16] Konasani V, Biswas M, Koleth PK. Healthcare Fraud Management using Big Data Analytics. An Unpublished Report by Trendwise Analytics, Bangalore, India. 2012.

[17] Rawte V, Anuradha G. Fraud Detection in Health Insurance using Data Mining Techniques. International Conference on Communication, Information & Computing Technology Jan. 16-17, 2015.

[18] Joudaki H, Rashidian A, Minaei-Bidgoli B, Mahmoodi M, Geraili B, Mahdi Nasiri M. Using Data Mining to Detect Health Care Fraud and Abuse. GJHS. 2015; 7(1).

[19] Woz´niak M, Grana M, Corchado E. A survey of multiple classifier systems as hybrid systems. IF. 2014; 16: 3–17.

J Health Inform Afr. 2017;4(1):32-41. DOI: 10.12856/JHIA-2017-v4-i1-161