Bank Fraud Detection using Community Detection Algorithm

Dhiman Sarma
Department of Computer Science and Engineering
Rangamati Science and Technology University
Rangamati, Bangladesh
dhiman001@yahoo.com

Wahidul Alam
Department of Computer Science and Engineering
University of Science and Technology Chittagong
Chittagong, Bangladesh
wahidulalam@ustc.ac.bd

Ishita Saha
Department of Computer Science and Engineering
University of Science and Technology Chittagong
Chittagong, Bangladesh
ishita_saha05@yahoo.com

Mohammad Nazmul Alam
Department of Computer Science and Engineering
Royal University of Dhaka
Dhaka, Bangladesh
mna235@yahoo.com

Mohammad Jahangir Alam
Department of Computer Science and Engineering
Southern University Bangladesh
Chittagong , Bangladesh
Jahangir@southern.edu.bd

Sohrab Hossain
Department of Computer Science and Engineering
East Delta University
Chittagong, Bangladesh
sohrab.h@eastdelta.edu.bd

***Abstract*- Bank fraud is a federal crime that involves fraudulent attempts aims for monetary gains by deceiving financial institutions. Every year, banks and financial institutions lose billions due to fraud. Fraudsters tempt bankers through scams to gain financial assets. The most common types of bank fraud include debit and credit card fraud, account fraud, insurance fraud, money laundering fraud, etc. Bankers are obliged to safeguard their financial assets as well as institutional integrity to armored the global financial system. Anti-fraud guard systems are regularly circumvented by fraudsters' dodging techniques. This paper proposed a system to detect bank fraud using a community detection algorithm that identifies the patterns that can lead to fraud occurrences. An agile method was used to design the web-based application to detect the fraud. The application functioned as a central hub between the banks and customers. Neo4j, a graph database, was used for creating and representing the database, and the Cypher query was used as a graph query language. The proposed system successfully detected all frauds presented during the test experiment. This paper will assist bankers to combat fraud by detecting and preventing similar occurrences.**

Keywords— Fraud Detection; Bank Fraud Detection; Community Detection Algorithm; Machine Learning; Neo4j; Cypher Query Language

I. INTRODUCTION

Banks and insurance companies are lucrative targets of fraudsters. They successfully seize billions worth financial assets every year. Fraud detection [1-4] plays an important role in minimizing these losses. Nevertheless, fraudsters continue their invasion by defeating all the existing and newly developed anti-fraudulent techniques with their clever dodgings. Among the effective tools, graph databases are used to identify fraud scenarios and other similar types of scams by the industry. It is also capable to stop fraud premises in real-time. Still, there has been no concrete system developed yet to prevent fraud. But the best practice is to develop methods for fraud detection. It can be achieved by using the individual data points and the connection among them. Sometimes these connections go unperceived until it is in the final stage. The connections among various data are important for identifying fraud because it can behold the best clues. Understanding the connection of various data and deriving meaning from these links can be drawn from one's existing data. The graph is used to rearrange the problem and to formulate a new look. Unlike most other ways of looking at data, the graph is designed to expose the connections. Graph databases are used to observe patterns very rapidly but they are difficult to detect using common symbolizes such as tables. At present, a large number of companies are using graph databases to solve different types of data related problems, including fraud detection.

In paper [5], researchers used data from the bank credit cardholders' dataset. The detection technique was developed based on the transactional data of credit card accounts.

The artificial neural network (ANN) was tested on various types of scam such as stolen credit card information, solicitation scam [6], camouflaged scam, and non-received issue scam. The ANN algorithm capable of detecting fraud more significantly with fewer false positives over fuzzy rule-based scam detection methods. The paper measured the network performance for this data set for accuracy and got the nearest to scam detection. The application was run on an

IBM 3080 computer and was recently in use for scam detection on that bank's data portfolio [7-12].

The objectives of these papers [13-15] were to develop a successful scam detection model to detect online banking scams and classification. Besides, the scope was to express its impact on the rapid and trustworthy detection of any "abnormal" transactions. These abnormal transactions were considered outliers for the banking system and treated as fraud activities.

Detecting communities play an important role in network science with multiple applications. Community detection is widely used in the area of machine learning applications. Because it helps to detect the presence of an internal network organization at a harsh level and detect further internal special affairs among the various nodes which are very hardtop to detect from direct empirical tests. Moreover, it allows researchers to comprehend the attributes of vigorous procedures taking place in a network. As syntagmatic examples, spreading processes of an outbreak are significantly affected by the community structure of the graph [16-19]. Data driven approaches are extremely useful to detect pattern in the dataset [20-23]. Security in banking sector can be achieved by using encryption and data driven based fraud detection system [24-27].

II. Methodology

The proposed model for detecting fraud is presented in figure 1. The proposed architecture in this paper consists of three major modules such as a training module, a storage module, and a detection module. The input data were entered by the legitimate bankers or users and captured by the training module. The training module passed test data to the storage module to store the data. The retrieval of data by the training module occurred when necessary. After then data was processed to the detection module to detect the fraud by using a cipher query builder and community system algorithm. The modules of figure 1 are described below:

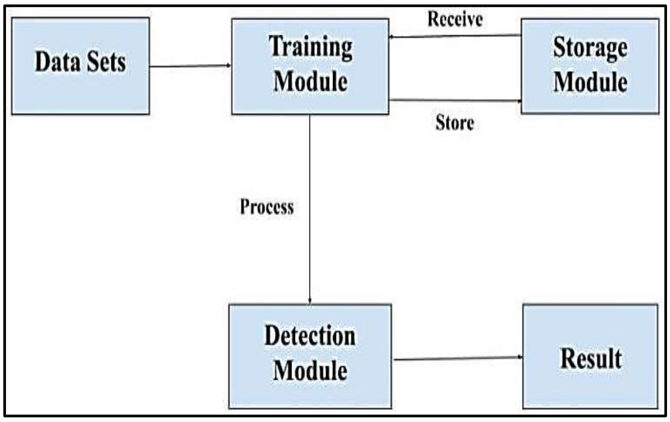


Fig. 1. Architectural of the proposed system

A. Training Module

A training module was used for making step-by-step processes to present a more realistic view of the information flow. The training module consists of raw data given by the users to detect if there are any fraudsters exist or not. NID, Passport number, Mobile Number, Address, etc. were being passed through the module and later, data were stored via the storage module, and finally processed by the detection module.

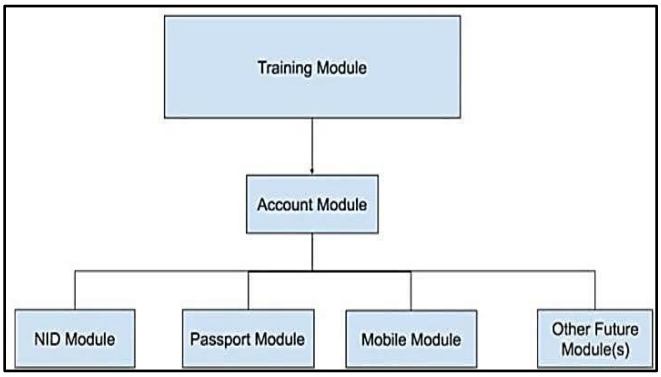


Fig. 2. Training module

B. Storage Module

The storage module is a system module to store the data from the training dataset. It stores the data in the system and sends the data to the training module while necessary. In this system, neo4j detailed with a bolt driver to store the data and reprocess the data to the training module if necessary. Besides, the neo4j cluster was being used to process the data in the storage module (figure 2).

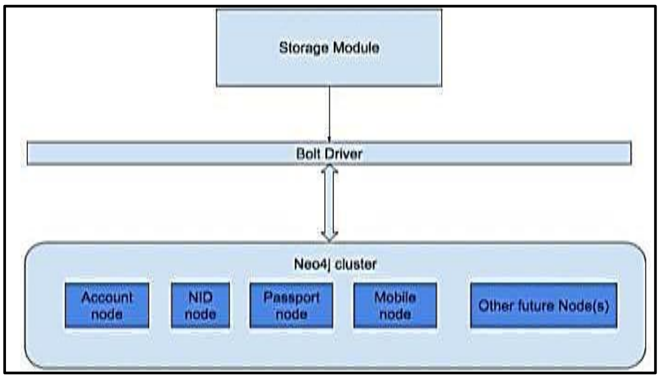


Fig. 3. Storage Module

C. Detection Module

Figure 4 shows the inside structure of the detection module. The core elements of the module are test data, cipher query builder, community detection algorithm, and response builder. Raw data were taken by the test data as a training dataset. The raw data were searched by the cypher query builder. Then the data passed through a community detection algorithm to check the connection between each node of the raw data and showed the result set via the response builder.

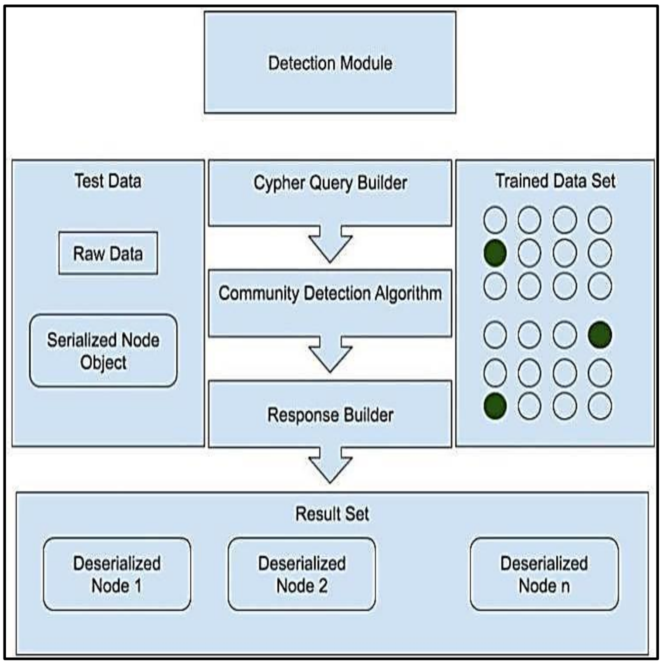
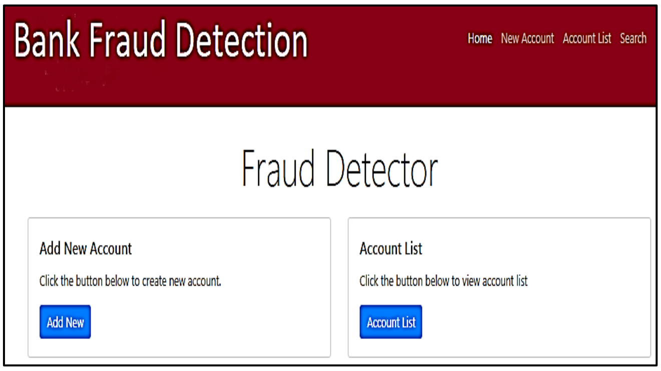


Fig. 4. Detection Module

V. RESULT AND DISCUSSION

A web-based application was developed to detect fraud in this paper. The application performs as a central hub between the users and the banks to easily detect fraud. The application was built as user friendly.


Bank Fraud Detection

Fraud Detector

Add New Account

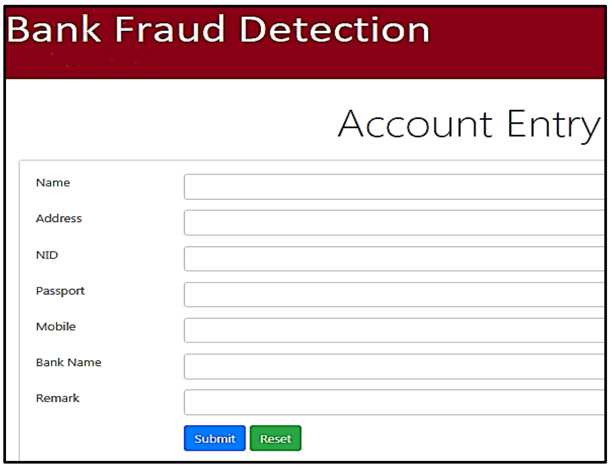
Click the button below to create new account.

Account List

Click the button below to view account list

Fig. 5. Home page.

Here, in figure 5, the data manager of every bank has the facility to input data of the account holder manually by clicking on the "*New Account*" button and the information of the clients are immediately stored in the database which could be viewed later by "*Account List*" button. But only the bank admins could have access to this panel to enter the input and visualize the list. This is the initial point where raw data were generated for the training module and storage module. Each time the data which were entered recently, can be retrieved from the storage module and viewed by the account list form (figure 6).


Bank Fraud Detection

Account Entry

Name

Address

NID

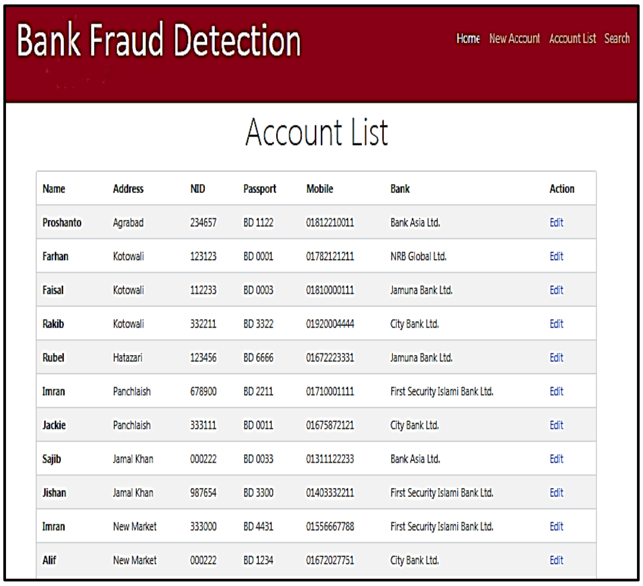
Passport

Mobile

Bank Name

Remark

Fig. 6. New Account.


Bank Fraud Detection

Account List

Name	Address	NID	Passport	Mobile	Bank	Action
Proshanto	Agrabad	234657	BD 1122	01812210011	Bank Asia Ltd.	Edit
Farhan	Kotowali	123123	BD 0001	01782121211	NRB Global Ltd.	Edit
Faisal	Kotowali	112233	BD 0003	01810000111	Jamuna Bank Ltd.	Edit
Rakib	Kotowali	332211	BD 3322	01920004444	City Bank Ltd.	Edit
Rubel	Hatazari	123456	BD 6666	01672223331	Jamuna Bank Ltd.	Edit
Imran	Panchlaish	678900	BD 2211	01710001111	First Security Islami Bank Ltd.	Edit
Jackie	Panchlaish	333111	BD 0011	01675872121	City Bank Ltd.	Edit
Sajib	Jamal Khan	000222	BD 0033	01311122233	Bank Asia Ltd.	Edit
Jishan	Jamal Khan	987654	BD 3300	01403332211	First Security Islami Bank Ltd.	Edit
Imran	New Market	333000	BD 4431	01556667788	First Security Islami Bank Ltd.	Edit
Alif	New Market	000222	BD 1234	01672027751	City Bank Ltd.	Edit

Fig. 7. Account List.

Later if any new customers want to open a new account you can check the data whether the information is valid or not by clicking on the *"Search"* button. There are three most important attributes provided which cannot be the same as two or more persons, these are national identification (NID) number, passport number, and mobile number.

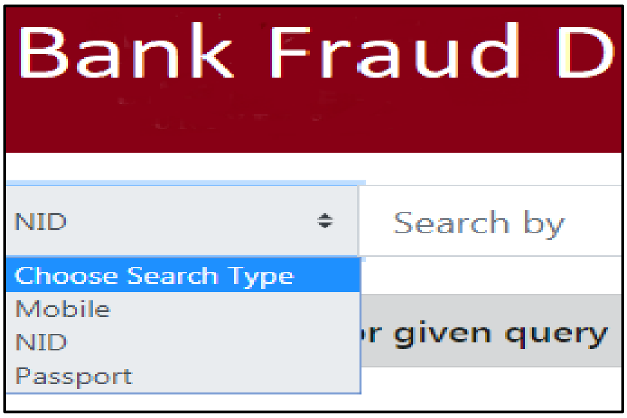
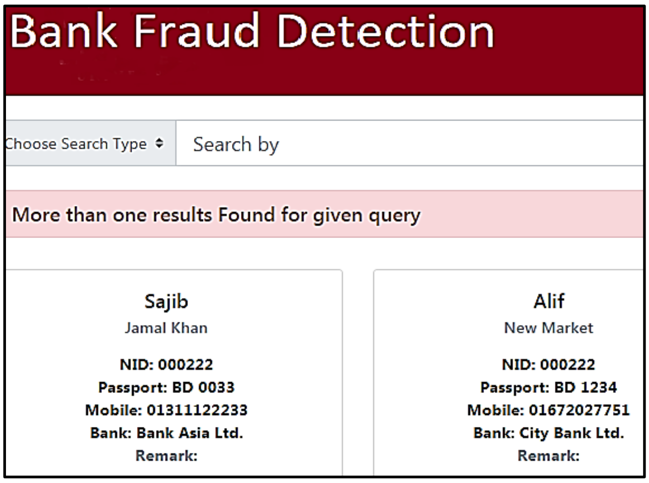

Bank Fraud D

Fig. 9. Search and the three most important attributes.

In case of any one of the attributes gets matched then it is detected immediately by the application and suspected as a fraud. Then Bank can initiate further investigation and take legal actions on them. Neo4j was used for creating and representing the database graphically as it makes it easier to detect by viewing the connections.



Bank Fraud Detection

More than one results Found for given query

Sajib
Jamal Khan
NID: 000222
Passport: BD 0033
Mobile: 01311122233
Bank: Bank Asia Ltd.
Remark:

Alif
New Market
NID: 000222
Passport: BD 1234
Mobile: 01672027751
Bank: City Bank Ltd.
Remark:

Fig. 11. For example two frauds of the same National Identification (NID) number.

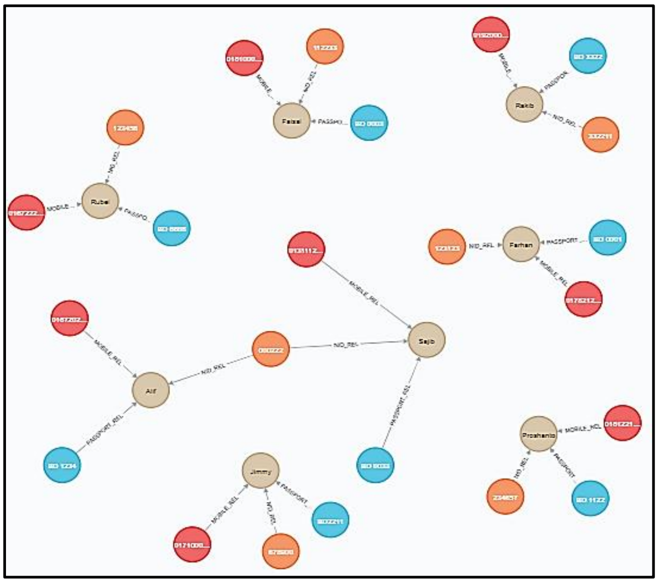
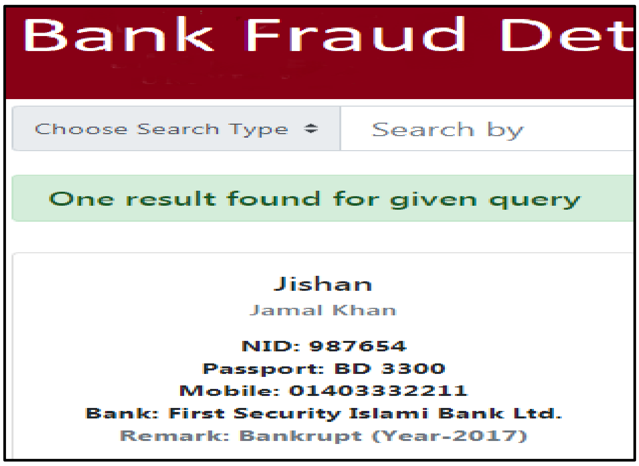


Fig. 8. Graphical representation of detecting fraud using Neo4j.

We tested another type of fraud in our experiment. Suppose, *"Jishan"*, who had an account in a bank named *'ABC Bank Limited'* two years ago and got bankrupt. This bankrupt information was updated in the bank's server. After two years later when he went to another bank to open a new account to apply for a new loan. Our system checked the attributes and found the previous details. It showed the bankrupt information and detected the event as fraud which restricted the application of *"Jishan"* to perform.



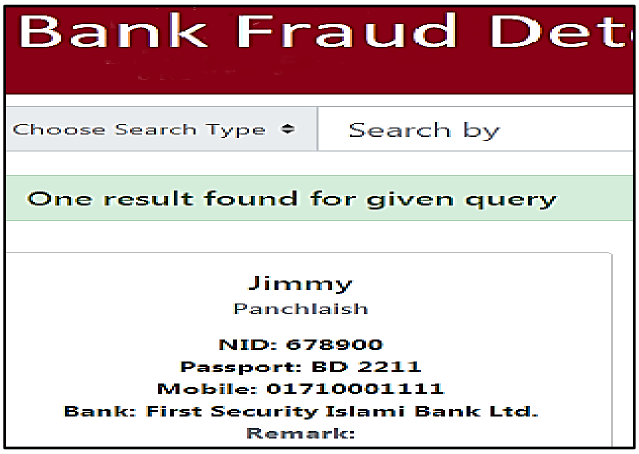
Bank Fraud Det

One result found for given query

Jishan
Jamal Khan
NID: 987654
Passport: BD 3300
Mobile: 01403332211
Bank: First Security Islami Bank Ltd.
Remark: Bankrupt (Year-2017)

Fig. 10. Fraud Scene.

In another scenario, if a new customer's data are always checked as valid and do not match or linked with any other account holder. Hence the customer's remarks pointed as clean.



Bank Fraud Det

One result found for given query

Jimmy
Panchlaish
NID: 678900
Passport: BD 2211
Mobile: 01710001111
Bank: First Security Islami Bank Ltd.
Remark:

Fig. 12. Clean account holder.

VI. Conclusion

Fraud detection is a challenging and intelligent task in the digital era. Expert systems and intelligent software are employed to defend fraud. This paper presented a fraud detection system using a community detection algorithm. It was developed through a web-based application that acts as an intermodal hub between the bankers and the customers. Neo4j, a graph database system, was implemented for searching and filtering fraud cases. Dataset, entered by the authorized banker, was passed through the training module to the storage module and can be retrieved by the training module on demand. These test data were then processed through a detection module where cypher query builder responded through the response builder and detect fraud. The test experiments were successfully detected by the application system and graphically represented to the users.

In the future, we will predict fraud associated with known fraudsters and their past activities by using a convolution neural network (CNN).

REFERENCES

[1] S. Daliri, "Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System," (in English), *Computational Intelligence and Neuroscience,* Article vol. 2020, p. 5, Feb 2020, Art. no. 6503459.

[2] S. M. Darwish, "An intelligent credit card fraud detection approach based on semantic fusion of two classifiers," (in English), *Soft Computing,* Article vol. 24, no. 2, pp. 1243-1253, Jan 2020.

[3] E. A. Minastireanu and G. Mesnita, "Methods of Handling Unbalanced Datasets in Credit Card Fraud Detection," (in English), *Brain-Broad Research in Artificial Intelligence and Neuroscience,* Article vol. 11, no. 1, pp. 131-143, Mar 2020.

[4] A. Somasundaram and S. Reddy, "Parallel and incremental credit card fraud detection model to handle concept drift and data imbalance," (in English), *Neural Computing & Applications,* Article vol. 31, pp. 3-14, Jan 2019.

[5] Y. Wu, Y. J. Xu, and J. Y. Li, "Feature construction for fraudulent credit card cash-out detection," (in English), *Decision Support Systems,* Article vol. 127, p. 11, Dec 2019, Art. no. 113155.

[6] S. Arora and M. P. S. Bhatia, "Fingerprint Spoofing Detection to Improve Customer Security in Mobile Financial Applications Using Deep Learning," (in English), *Arabian Journal for Science and Engineering,* Article vol. 45, no. 4, pp. 2847-2863, Apr 2020.

[7] O. Ata and L. Hazim, "Comparative Analysis of Different Distributions Dataset by Using Data Mining Techniques on Credit Card Fraud Detection," (in English), *Tehnicki Vjesnik-Technical Gazette,* Article vol. 27, no. 2, pp. 618-626, Apr 2020.

[8] A. Eshghi and M. Kargari, "Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty," (in English), *Expert Systems with Applications,* Article vol. 121, pp. 382-392, May 2019.

[9] R. Seidlova, J. Pozivil, and J. Seidl, "Marketing and business intelligence with help of ant colony algorithm," (in English), *Journal of Strategic Marketing,* Article vol. 27, no. 5, pp. 451-463, 2019.

[10] X. L. Xu, N. Hu, M. Trovati, J. Ray, F. Palmieri, and H. M. Pandey, "DLCD-CCE: A Local Community Detection Algorithm for Complex IoT Networks," (in English), *Ieee Internet of Things Journal,* Article vol. 7, no. 5, pp. 4607-4615, May 2020.

[11] A. Zamir *et al.*, "Phishing web site detection using diverse machine learning algorithms," (in English), *Electronic Library,* Article vol. 38, no. 1, pp. 65-80, Jan 2020.

[12] Z. K. Zandian and M. R. Keyvanpour, "Feature Extraction Method Based on Social Network Analysis," (in English), *Applied Artificial Intelligence,* Article vol. 33, no. 8, pp. 669-688, Jul 2019.

[13] M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), *Intelligent Decision Technologies-Netherlands,* Article vol. 13, no. 2, pp. 229-270, 2019.

[14] F. Karimi, S. Lotfi, and H. Izadkhah, "Multiplex community detection in complex networks using an evolutionary approach," (in English), *Expert Systems with Applications,* Article vol. 146, p. 20, May 2020, Art. no. 113184.

[15] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," (in English), *Information Sciences,* Article vol. 479, pp. 448-455, Apr 2019.

[16] R. B. Xu, Y. Che, X. M. Wang, J. X. Hu, and Y. Xie, "Stacked autoencoder-based community detection method via an ensemble clustering framework," (in English), *Information Sciences,* Article vol. 526, pp. 151-165, Jul 2020.

[17] I. Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), *Information Sciences,* Article vol. 485, pp. 319-346, Jun 2019.

[18] M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), *Financial and Credit Activity-Problems of Theory and Practice,* Article vol. 1, no. 28, pp. 36-45, 2019.

[19] G. S. Carnivali, A. B. Vieira, A. Ziviani, and P. A. A. Esquef, "CoVeC: Coarse-grained vertex clustering for efficient community detection in sparse complex networks," (in English), *Information Sciences,* Article vol. 522, pp. 180-192, Jun 2020.

[20] F. Ahmed, et al., "A Combined Belief Rule based Expert System to Predict Coronary Artery Disease," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 252-257.

[21] S. Hossain, et al., "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-6.

[22] S. Hossain et al. "A Critical Comparison between Distributed Database Approach and Data Warehousing Approach." International Journal of Scientific & Engineering Research, Article 5.1 (2014): 196-201.

[23] K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.

[24] D Sarma., 2012. Security of Hard Disk Encryption. Masters Thesis. Identifiers: urn:nbn:se:kth:diva-98673(URN), Royal Institute of Technology.

[25] S. Hossain, D. Sarma, R. J. Chakma, W. Alam, M. M. Hoque, and I. H. Sarker, "A Rule-Based Expert System to Assess Coronary Artery Disease Under Uncertainty," in Computing Science, Communication and Security, Singapore, 2020, pp. 143-159: Springer Singapore.

[26] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[27] H. Alqahtani et al., "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, 2020, pp. 121-131: Springer Singapore.