iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

FraudAuditor: A Visual Analytics Approach for Detecting Collusive Fraud in Health Insurance Systems

Mr. Thathineni Jagadeesh[1], Penke Vamsi Krishna[2], Medisetti Manasa[3], Puppala Sri Vinay[4], Khandavalli Ajay[5], Katta Akash[6]

[1] *Assistant Professor, Department of Computer Science and Engineering (AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

[2,3,4,5,6] *B.Tech Students, Department of Computer Science and Engineering(AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

Abstract: ***Health insurance fraud, particularly collusive fraud involving coordinated actions between patients and healthcare providers, poses a significant challenge to modern healthcare systems, leading to substantial financial losses and inefficient resource utilization. Traditional fraud detection methods, including rule-based systems and supervised machine learning models, often fail to identify such organized fraud patterns due to the lack of labeled datasets and the complexity of relational behaviors. To address these limitations, this paper proposes FraudAuditor, an unsupervised graph-based visual analytics framework designed to detect and analyze collusive fraud without relying on pre-labeled data. The system constructs a Weighted Co-Visit Network, where patients are represented as nodes and edges indicate shared provider visits within temporal windows. The strength of relationships is quantified using synchronization, frequency, and contextual similarity metrics. The Louvain community detection algorithm is applied to identify densely connected clusters that may represent fraudulent groups. Furthermore, an interactive dashboard incorporating network visualization, PCA-based projections, and multi-dimensional risk profiling enhances interpretability and supports effective investigation. Experimental evaluation demonstrates that the proposed system improves detection capability, reduces false positives, and provides a scalable and explainable solution for healthcare fraud analysis.***

Keywords: Health Insurance Fraud, Collusive Fraud Detection, Graph-Based Analysis, Louvain Community Detection, Visual Analytics, Weighted Co-Visit Network, Unsupervised Learning.

I. INTRODUCTION

Health insurance fraud is a critical issue that significantly impacts healthcare systems worldwide by causing financial losses, increasing insurance premiums, and reducing access to genuine medical services. Among various fraud types, collusive fraud is particularly difficult to detect as it involves coordinated activities between multiple patients and healthcare providers. These fraud schemes are designed to mimic legitimate medical behavior, making them difficult to identify using traditional detection techniques.

Conventional fraud detection systems rely heavily on rule-based approaches or supervised machine learning models, which primarily focus on identifying individual anomalies such as high claim amounts or frequent visits. However, these approaches fail to capture the structural relationships and coordinated behavior among multiple entities, which are key indicators of collusive fraud . Additionally, the lack of reliable labeled datasets and evolving fraud strategies further limits the effectiveness of these methods.

To overcome these challenges, this work introduces FraudAuditor, a novel unsupervised framework that focuses on relationship-based fraud detection. By modeling healthcare data as a graph structure and applying community detection techniques, the system identifies hidden patterns of coordination that are not visible in traditional tabular analysis. Furthermore, the integration of visual analytics enables human-in-the-loop investigation, improving interpretability and decision-making.

A. Problem Statement

Existing fraud detection systems face challenges such as lack of labeled data, high false positive rates, and inability to detect coordinated fraud patterns. These systems often fail to distinguish between legitimate medical behavior and collusive fraud. Therefore, there is a need for a robust, unsupervised, and explainable system capable of identifying complex fraud structures.

B. Motivation

The motivation behind this work stems from the increasing financial losses caused by healthcare fraud and the limitations of existing detection systems. There is a strong need for intelligent systems that can detect hidden fraud patterns while providing clear and interpretable insights to auditors.

C. Scope

The scope of this project includes data ingestion, preprocessing, graph construction, community detection, risk profiling, and visualization of healthcare claim data. The system is capable of handling large datasets and provides interactive dashboards for detailed investigation and analysis of suspicious activities.

II. LITERATURE SURVEY

The field of fraud detection has evolved significantly from rule-based systems to advanced data-driven and graph-based analytical approaches. Early systems relied on predefined thresholds and expert rules, which were effective for detecting simple anomalies but lacked adaptability to complex fraud patterns. With the advancement of machine learning, researchers introduced supervised and unsupervised techniques to improve detection accuracy. However, supervised models require large labeled datasets, which are often unavailable in healthcare fraud scenarios.

Recent studies emphasize the importance of graph-based analysis and social network modeling for detecting collusive fraud. These approaches focus on analyzing relationships between entities rather than treating them independently. Techniques such as bipartite graph modeling, community detection, and temporal analysis have proven effective in identifying coordinated fraud patterns . The Louvain algorithm, in particular, has gained popularity due to its efficiency in detecting communities in large-scale networks.

Additionally, the integration of visual analytics has enhanced fraud detection by enabling human experts to interpret complex patterns through interactive visualizations. This combination of automated detection and human interpretation improves accuracy, reduces false positives, and supports informed decision-making.

S.No	Author(s) & Year	Methodology	Dataset	Key Contribution	Limitations
1	Bolton & Hand (2002)	Statistical Fraud Detection	Financial Data	Early fraud detection techniques	Limited scalability
2	Phua et al. (2010)	Data Mining Approaches	Insurance Data	Survey on fraud detection methods	Lacks real-time analysis
3	Akoglu et al. (2015)	Graph-based Anomaly Detection	Network Data	Introduced graph-based fraud detection	Computational complexity
4	Pandit et al. (2007)	NetProbe Algorithm	Auction Networks	Detects fraud using network analysis	Limited domain applicability
5	Beutel et al. (2013)	CopyCatch Graph Method	Social Networks	Detects synchronized behavior	Requires large datasets
6	Kou et al. (2004)	Machine Learning Models	Financial Data	ML-based fraud detection framework	Needs labeled data
7	Fortunato (2010)	Community Detection Survey	Network Data	Overview of clustering methods	No specific fraud focus
8	Blondel et al. (2008)	Louvain Algorithm	Large Networks	Efficient community detection	May miss small clusters
9	Weber et al. (2018)	Visual Analytics	Financial Fraud	Combines visualization with ML	User-dependent analysis
10	Van Vlasselaer et al. (2017)	Collective Fraud Detection	Insurance Data	Detects group fraud patterns	Complex implementation
11	Jiang et al. (2016)	Graph Neural Networks	Transaction Data	Advanced deep learning models	High computational cost
12	Ribeiro et al. (2020)	Explainable AI	Fraud Systems	Improves model interpretability	Trade-off with accuracy

III. BACKGROUND WORK

A. Graph-Based Fraud Detection in Healthcare Systems

Traditional fraud detection systems in healthcare primarily rely on rule-based or statistical approaches that focus on identifying anomalies at the individual level. However, these methods fail to capture the relational dependencies that characterize collusive fraud. Graph-based modeling has emerged as an effective alternative, where entities such as patients and healthcare providers are represented as nodes, and their interactions form edges. This representation enables the identification of hidden patterns and relationships within large datasets. By analyzing connectivity and interaction structures, graph-based approaches provide deeper insights into coordinated fraudulent activities that remain undetected in conventional tabular analysis.

B. Community Detection Techniques for Collusive Behavior Analysis

Community detection plays a vital role in identifying clusters of entities that exhibit strong internal connections. In fraud detection, these clusters often correspond to coordinated groups involved in fraudulent activities. Among various algorithms, the **Louvain method** is widely adopted due to its efficiency and scalability in handling large networks. It partitions the graph by maximizing modularity, thereby identifying densely connected communities. In the context of healthcare fraud, patients who frequently visit the same providers within short time intervals form tightly connected clusters, indicating potential collusion. This approach eliminates the dependency on labeled datasets and enables the detection of emerging fraud patterns.

C. Visual Analytics for Fraud Investigation

While automated detection methods are effective, their lack of interpretability often limits their practical usability. Visual analytics bridges this gap by integrating data visualization with analytical processing, enabling human experts to interpret complex patterns. Interactive dashboards, network graphs, and multi-dimensional visualizations allow auditors to explore relationships, identify suspicious clusters, and validate fraud cases effectively. Techniques such as force-directed graph layouts and PCA-based projections enhance the understanding of structural and behavioral patterns. This human-in-the-loop approach improves decision-making, reduces false positives, and increases trust in the system.

IV. PROPOSED MODEL

A. Overview of the Proposed System

The proposed system, **FraudAuditor**, is an unsupervised visual analytics framework designed to detect collusive fraud in health insurance datasets. Unlike traditional methods, the system focuses on **relationship-centric modeling** by transforming structured healthcare data into a graph representation. It integrates graph construction, community detection, risk profiling, and visualization into a unified pipeline, enabling both automated detection and human-assisted investigation.

B. Weighted Co-Visit Network Construction

The system constructs a **Weighted Co-Visit Network**, where each node represents a patient, and edges are formed between patients who visit the same healthcare provider within a defined temporal window. The strength of each edge is calculated using a composite weighting function that incorporates:

- Visit frequency
- Temporal proximity
- Contextual similarity (diagnosis, treatment, claim amount)

This weighted representation ensures that stronger relationships reflect higher levels of coordination, enabling the detection of suspicious patterns.

C. Unsupervised Community Detection Using Louvain Algorithm

To identify potential fraud groups, the system applies the **Louvain community detection algorithm** on the constructed network. This algorithm partitions the graph into communities by maximizing modularity, grouping together patients with dense internal connections. These communities represent potential collusive fraud rings, as they exhibit synchronized behavior patterns. Since this approach does not rely on labeled data, it remains adaptable to new and evolving fraud strategies.

D. Risk Profiling and Analytical Scoring

The system incorporates a multi-dimensional risk profiling mechanism to evaluate detected communities and individual patients. Risk scores are computed based on:

- Financial indicators (claim amounts, billing frequency)
- Temporal synchronization (visit overlap patterns)
- Clinical similarity (diagnosis and treatment patterns)

These metrics are combined to generate a comprehensive risk score, enabling prioritization of high-risk cases for investigation.

E. Visual Analytics and Investigation Dashboard

FraudAuditor integrates an interactive dashboard that supports detailed exploration of detected fraud patterns. Key features include:

- Force-directed network visualization of patient communities
- PCA-based projection for cluster separation analysis
- Risk radar charts for multi-dimensional profiling
- Contextual side panels for patient-level details

This visualization layer enhances interpretability and allows auditors to validate algorithmic findings effectively.

F. System Workflow

The overall workflow of the proposed model is as follows:

- Upload structured healthcare claim datasets.
- Perform data preprocessing and validation.
- Construct the Weighted Co-Visit Network.
- Apply Louvain algorithm for community detection.
- Compute risk scores for identified communities.
- Visualize results through an interactive dashboard.
- Enable auditors to investigate and validate suspicious cases.

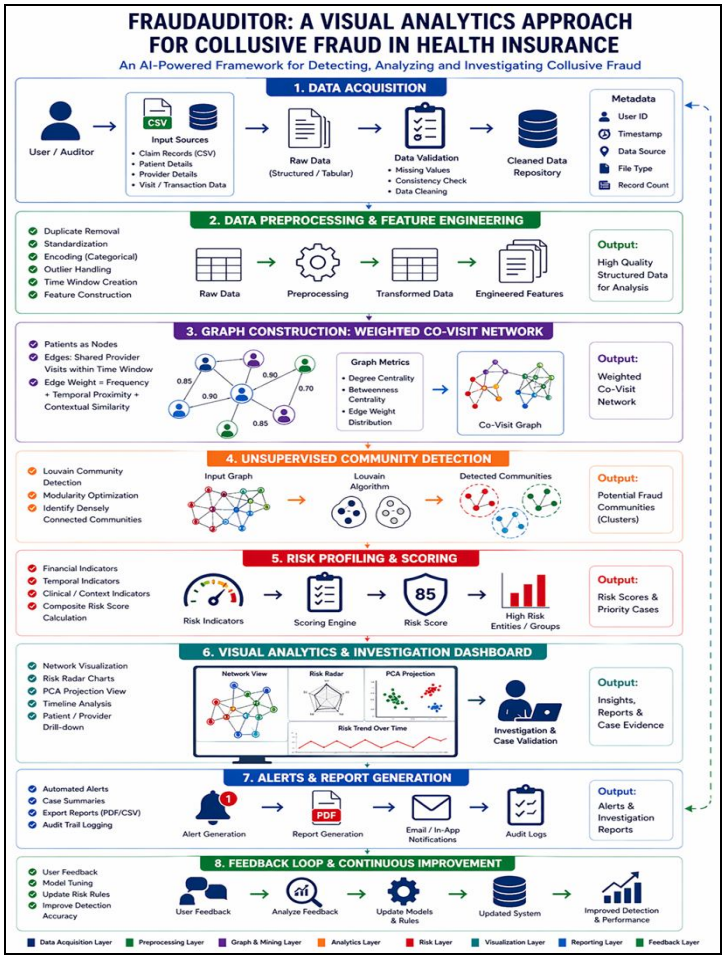


Figure 1 illustrates the Proposed architecture

V. IMPLEMENTATION RESULTS

The experimental setup of the proposed FraudAuditor system is implemented using Python, leveraging libraries such as NetworkX for graph construction, Pandas for data preprocessing, and Scikit-learn for analytical computations. The system is evaluated on healthcare insurance datasets containing patient records, provider details, and transaction/visit histories collected from simulated and publicly available sources. Data preprocessing includes cleaning, handling missing values, normalization, and constructing temporal windows for co-visit analysis. A weighted co-visit network is generated where edge weights are computed based on visit frequency, temporal proximity, and contextual similarity. The Louvain community detection algorithm is applied to identify potential collusive groups, followed by multi-dimensional risk scoring using financial, temporal, and clinical indicators. The performance of the system is assessed using metrics such as detection accuracy, precision, recall, and reduction in false positives, along with qualitative evaluation through visual analytics dashboards. Experiments are conducted on a system with adequate computational resources to ensure scalability and efficient processing of large healthcare datasets.

1) Home Page

Figure 2. Enter your credentials and click Login.

Figure 2 represent the Enter your credentials and click Login.

2) Overview of the System / Click Launch Dashboard

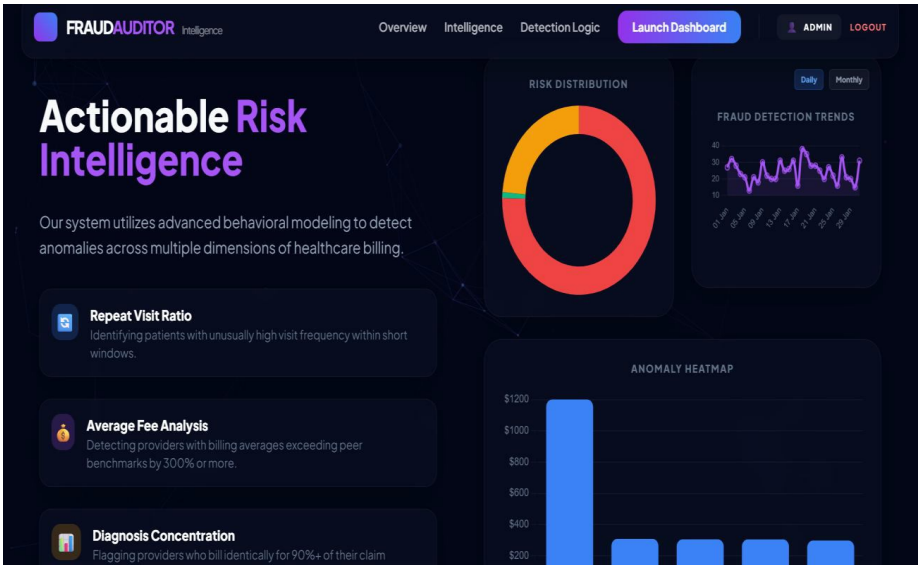


Figure 3. Overview of the System / Click Launch Dashboard

3) Dashboard Section of Expense Analysis Tracker

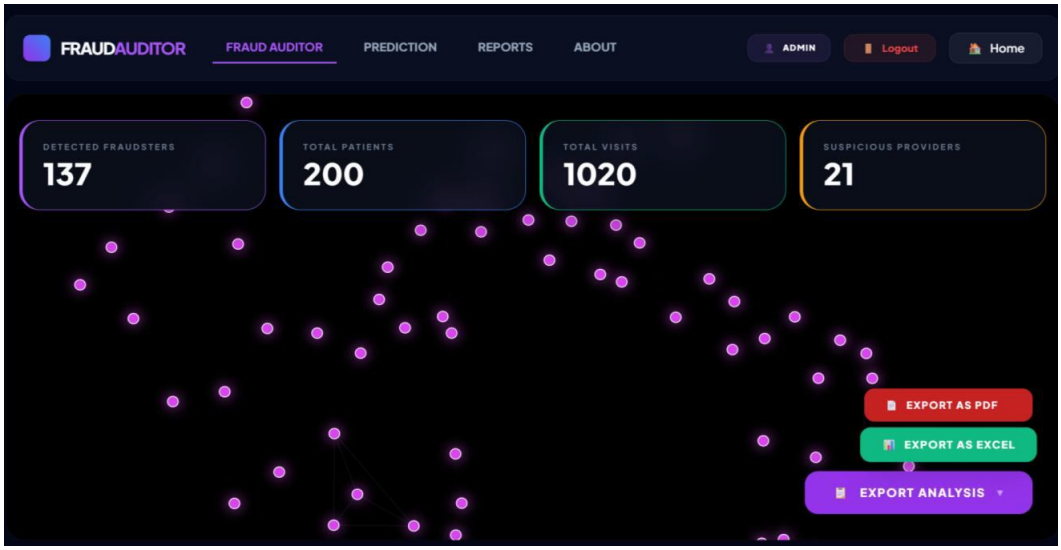


Figure 4. Click on "Export Analysis" to download the train data report

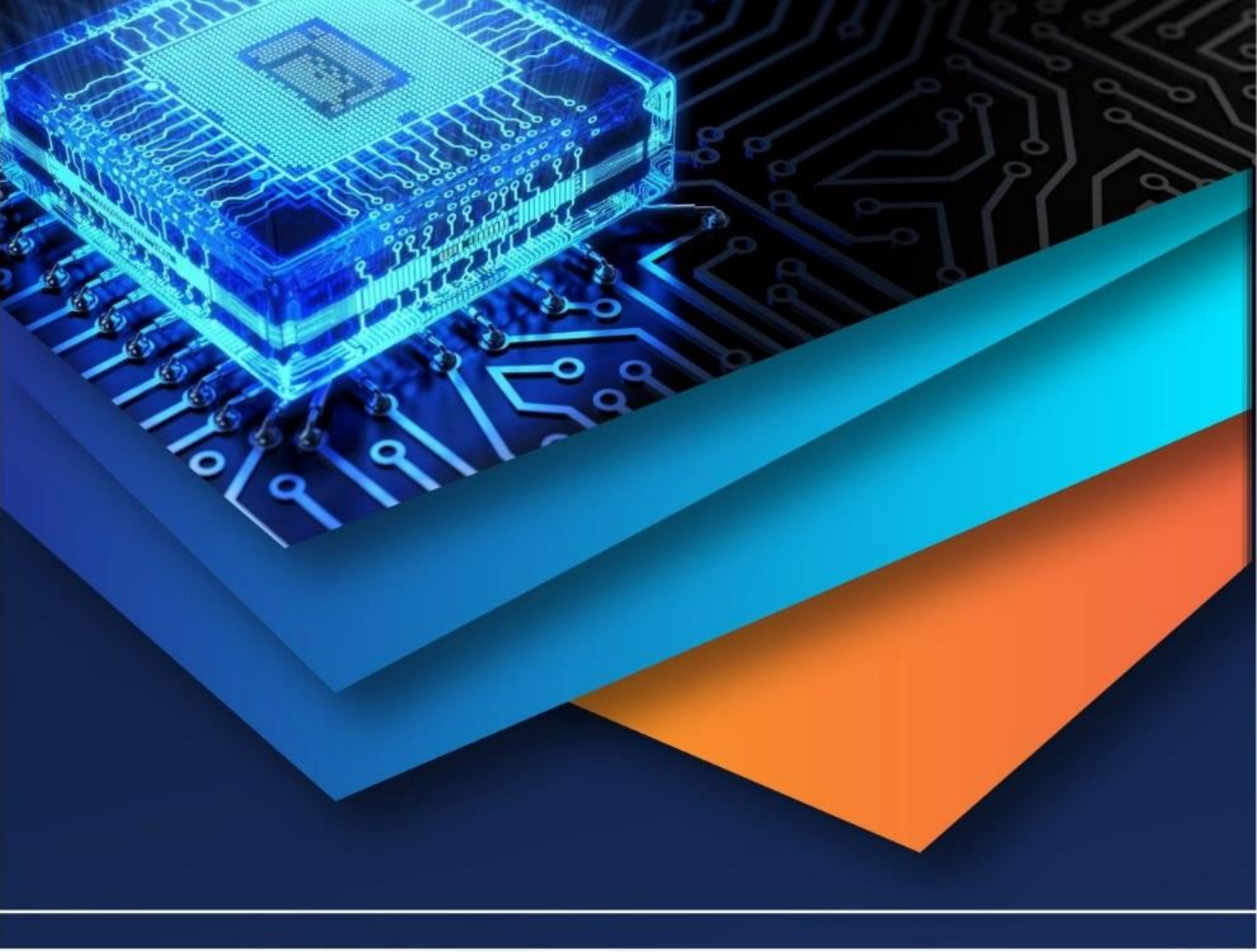
Figure 4 illustrates click on "Export Analysis" to download the train data report

VI. CONCLUSION

This paper presented FraudAuditor, a visual analytics-driven framework for detecting collusive fraud in health insurance systems. Unlike traditional rule-based and supervised approaches, the proposed model adopts an unsupervised, graph-based methodology that focuses on relationships between entities rather than isolated transactions. By constructing a weighted co-visit network and applying the Louvain community detection algorithm, the system effectively identifies densely connected groups that may represent coordinated fraudulent activities.The integration of multi-dimensional risk profiling further enhances detection by combining financial, temporal, and contextual indicators to prioritize suspicious cases. In addition, the incorporation of interactive visual analytics enables auditors to interpret complex patterns, validate findings, and make informed decisions with greater transparency. Experimental observations indicate that the proposed approach improves fraud detection capability, reduces false positives, and scales efficiently for large healthcare datasets.Overall, FraudAuditor provides a robust, explainable, and scalable solution for addressing the growing challenge of collusive fraud in health insurance. Future work can extend this framework by incorporating graph neural networks, real-time streaming data analysis, and advanced explainable AI techniques to further enhance detection accuracy and adaptability.

REFERENCES

[1] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," Statistical Science, vol. 17, no. 3, pp. 235–255, 2002.
[2] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," arXiv preprint arXiv:1009.6119, 2010.
[3] L. Akoglu, H. Tong, and D. Koutra, "Graph-based anomaly detection and description: A survey," Data Mining and Knowledge Discovery, vol. 29, no. 3, pp. 626–688, 2015.
[4] S. Pandit, D. Chau, S. Wang, and C. Faloutsos, "NetProbe: A fast and scalable system for fraud detection in online auction networks," in Proc. WWW, 2007, pp. 201–210.
[5] A. Beutel et al., "CopyCatch: Stopping group attacks by spotting lockstep behavior in social networks," in Proc. WWW, 2013, pp. 119–130.
[6] Y. Kou, C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang, "Survey of fraud detection techniques," in Proc. IEEE Int. Conf. Networking, Sensing and Control, 2004, pp. 749–754.
[7] S. Fortunato, "Community detection in graphs," Physics Reports, vol. 486, no. 3–5, pp. 75–174, 2010.
[8] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics, vol. 2008, no. 10, p. P10008, 2008.
[9] M. Weber, M. H. Lee, and K. Böhm, "Visual analytics for fraud detection: A survey," IEEE Trans. Visualization and Computer Graphics, vol. 24, no. 1, pp. 1–20, 2018.
[10] V. Van Vlasselaer et al., "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," Decision Support Systems, vol. 75, pp. 38–48, 2015.
[11] W. Jiang, W. Cao, and Z. Liu, "Graph neural networks for fraud detection: A review," IEEE Access, vol. 8, pp. 156793–156805, 2020.
[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016, pp. 1135–1144.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY