International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Anomaly Detection in Healthcare Data Warehouses Using ML

Arun Kumar Ramachandran Sumangala Devi

Architect II- Software Testing UST Global Inc, Glen Allen, Virginia, USA

Abstract:
Fast emergence of digitalization leads to unprecedented growth in the generation of the healthcare sector-particularly EHRs and medical equipment data. This extended the way for challenges for integrity in managing data and anomaly detection, including fraudulent transactions, medication errors, and many more system failures. Modern healthcare data poses a challenge to traditional methods of anomaly detection due to high and complex dimensionality. Machine learning provides a strong solution, using algorithms such as Gaussian Mixture Models, One-Class SVM and deep learning algorithms such as Autoencoders, and Recurrent Neural Networks in the detection of anomalies in healthcare data warehouse settings [1]. This study reports how ML can help advance care for patients, enable the validity of the data and reduce costs through real-time monitoring, fraud detection, and early detection of diseases. Applying anomaly detection through ML would most likely bring better operational performance, patient safety, and decision-making in health care for organizations as issues of poor data quality, lack of interpretability of models, and real-time detection would be addressed [2].

Keywords: Machine learning, fraud detection healthcare data warehouses, anomaly detection, patient safety, operational efficiency.

1. Introduction

The present landscape of digitalisation in the healthcare sector encompasses technologies such as machine learning for detecting anomalies within the healthcare data warehousing process. The emerging growth of healthcare data delivers various opportunities as well as possible issues. It has become intensively challenging for the healthcare sector to address anomalies and ensure data integrity as a huge number of data has been generated from electronic health records [3]. Issues and concerns raised in the health care industry showed that there was the possibility of facing some of the risks as fraud, medical mistakes and system failure [4]. Acceptable procedures occasionally miss the abnormalities since the apprehensive healthcare information grows vast and higher-dimensional today. Contemporary technologies like machine learning can be used for detections of anomalies in healthcare data to detect unusual patterns that can lead to critical problems [5]. It is suggested that Machine Learning (ML) should be applied to standardise the means to perform anomaly detection in healthcare data warehouses presents the most suitable solution. The support of intricate and sophisticated ML algorithms enables the healthcare industry comprehend vast data. It also helps them filter some kinds of patterns and signals of development of specific problems that may occur. The knowledge on the subject was expanded through this research by exploring ML techniques as a means of improving the anomaly detection capability in retrieving

healthcare data warehouses. This further aims to enhance patients' treatment, data accuracy and ultimately the costs on the same.

2. Solution

Application of advanced types of Machine Learning helps in detection of anomalies in the healthcare data marts. The method of acquiring the data from one or even multiple sources including: EHR; medical images; IoT and many more help to address the challenges in data management [6]. But it is also crucial to preprocess the inputs given to these machines and cleaning and normalizing form part of the preprocessing. The machine provides a strong approach to the detection of anomalies in healthcare data warehouses. Through the analysis of the broad data sets, ML algorithms were able to address the patterns and variations which highlighted the fraudulent activities and safety concerns.

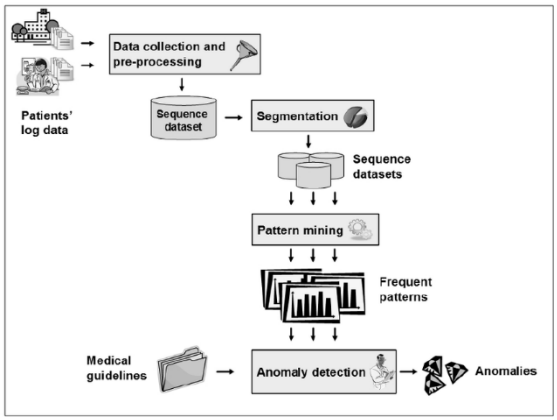

Figure 1: Anomaly detection framework in healthcare system [6]

The enhancing availability of electronic medical scores makes it feasible to reconstruct the patient treatment. Hence, developing methods to predict deviations through the pattern could help determine the management. Given the high dimensional settings, statistical methods such as Gaussian Mixture Models and One-Class SVM are suitable for the detection of anomalies within the high-dimensional data [7]. Such algorithms are typically able to pick out data points that deviate strongly from the bulk. Deep Learning techniques include the Recurrent and Autoencoders Neural networks which are intensely powerful over the time-series data [8]. Autoencoders learn to reconstruct normal patterns, and deviations from those patterns may be identified as anomalies. RNNs enable the model to capture the temporal dependencies within the data which is one of the reasons for its effectiveness. The choice of the ML algorithm would depend on the different characteristics of the healthcare data including the dimensionality and distribution of the data. The relevant algorithms have to be assessed with caution to decide the right application for the given use cases.

The models are then implemented in the healthcare data warehouse setting. Real-time data monitoring is adopted whereby data that are received are analysed in real-time to identify exceptions. The outcome extracted by the algorithm is visualized using heat maps and time series plots to enable the healthcare provider to comprehend the outcome and take the subsequent appropriate action.

3. Applications of the Solution

The machine learning techniques are used for anomaly detection within the healthcare data for addressing the abnormal pace and patterns of the essential signs which may suggest the issues in health. The possible applications are as follows:

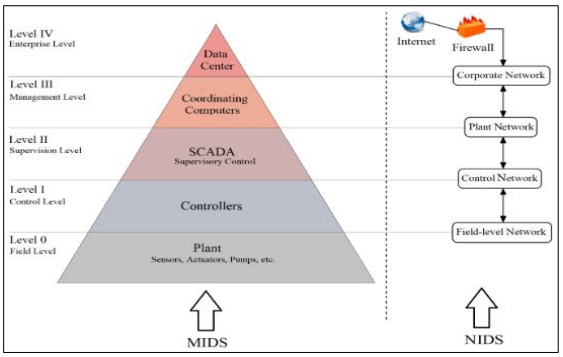

Figure 2: Anomaly detection architecture [8]

The above figure showing the anomaly detection architecture which could beneficial for monitoring the network traffic. It would effectively investigate the doubtful activities within system measurement data.

Patient monitoring: ML technology assists in the analysis of data from wearable devices and includes medical monitoring within the healthcare sector. This helps to detect the usual patterns and rate of heartbeat, oxygen level and blood pressure. Anomalies in medical devices and patient monitoring systems can be seen as potential safety risks [9]. The detection of these anomalies in the real world can guide healthcare providers to take immediate actions through the prevention of negative effects and ensuring patient safety. Anomaly detection also assists in averting the outbreak of disease along with infection in a population.

Fraud detection: Healthcare fraud is considered to be a huge problem running into billions of dollars yearly. ML techniques help healthcare organisations to identify insurance fraud so that action can be taken timely [10]. It can best be detected using techniques of anomaly detection, especially when an area in the billing process has some original or suspicious insurance claims; such areas will be addressed appropriately and financial loss avoided in the course of ensuring fair healthcare practices.

Disease detection: The machine learning algorithms help to identify the anomalies in the MRI scans and X-rays to indicate major diseases such as cancer. The machine learning models help to analyse the real-time data for the detection of unusual patterns of the patient's vital signs [11]. For instance, the unexpected change in the health beat could be a signal for cardiac arrest. In addition to this, in the medical setting, the identification of anomalies in MRI scans and X-rays suggests the indication of severe disease such as cancer which can be identified earlier than the human physicians through this technology.

Quality control: Machine learning technologies assist in the identification of errors in medical information, malfunctions in the equipment and inconsistencies in the treatment protocols [12]. This further improves patient safety and reduces errors in the medical data.

4. Benefits of the Solution

The successful incorporation of ML technology for anomaly detection in the healthcare data warehouse provides significant opportunities.

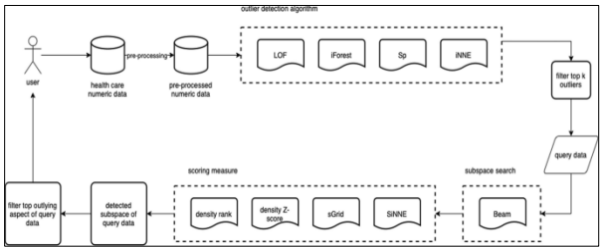

Figure 3: Detection of anomalies in healthcare data [10]

the above figure showing the process flow of anomalies detection in healthcare data. It showing the whole process of anomalies detection within the healthcare data management. However, the advantage of this system is given in below.

Improved patient outcomes: Such problems can be identified early, and corrections made before they cause complications that could harm the patient, hence promoting patient safety.

Reduced costs: There are numerous ways that fraud, wasteful resources and equipment downtimes can be identified in healthcare organizations [13]. This provides methods that will help to reduce the expenditure in the long term.

Enhanced operational efficiency: Depending on these dimensions, anomaly detection enables one to improve efficiency, reduce non-conformities and determine new solutions which would create more efficiency [14].

Increased patient safety: Through the identification of the errors regarding medical records and treatments some medical errors can be stopped thereby protecting the lives of patients.

Data-driven decision-making: Additionally, an approach to anomaly detection gives benefits for analysing a large amount of healthcare data and makes it possible to make decisions based on the analysis of such data [15].

Though anomaly detection in healthcare data provides significant potential, it also presents different challenges.

Data Quality and Availability: Often the data in healthcare may be missing parts or values which can make it difficult to discern the anomalous points. Ensuring the existence and quality, are two vital factors in anomaly detection [16].

Interpretability: The anomaly detection algorithms themselves usually constitute the bare black box model and are hard to explain. A large concern is needed for healthcare practitioners to trust anomaly detection models; interpretable models are ideal for this goal.

Real-Time Detection: In any case, relying on real-time data for anomaly detection plays a crucial role in the timely making of a decision in healthcare. Currently, it is difficult to find efficient algorithms that would allow the distribution of a significant amount of information to a large number of users in real-time. Despite these issues, the future of anomaly detection in health-care data stands as quite promising. Advancements in machine learning, integration of diversified sources of health care and improvement techniques in data quality data can further enhance the accuracy and efficiency of the anomaly-detection systems.

5. Conclusion

The integration of ML systems in healthcare data warehouses fosters data integrity within healthcare organisations which underlines the essence of this research. This research underlines that using ML provides a vigorous approach to identifying abnormalities in healthcare data repositories. Through the incorporation of ML algorithms, healthcare organizations may learn more about the data they collect, as well as the risks and opportunities for the patients. The sophistication of these ML techniques is set to improve the healthcare setting meaning the future holds complexities in this healthcare aspect as it tweaks the healthcare operations to suit patient requirements. This research concludes that the utilization of ML in anomaly detection of healthcare data warehouses can solve problems arising due to the complexity and volume of healthcare data quite robustly. ML thereby automates the detection of anomalies, enhancing patient safety, operational efficiency, and data integrity. Its applications are wide-ranging: early detection

of diseases, fraud prevention, and regulatory compliance. The ever-growing amount of healthcare data will mean that the integrity of healthcare systems can be ensured and perfected which underscores the significance of integrating ML-based anomaly detection in healthcare data warehouses.

References

1. Al-amri, R. et al., "A Review of Machine Learning and Deep Learning Techniques for Anomaly Detection in IoT Data," Applied Sciences, vol. 11, no. 12, p. 5320, 2021.
2. Gao, B. et al., "Enhancing Anomaly Detection Accuracy and Interpretability in Low-Quality and Class Imbalanced Data: A Comprehensive Approach," Applied Energy, vol. 353, p. 122157, 2024.
3. Tayefi, M. et al., "Challenges and Opportunities Beyond Structured Data in the Analysis of Electronic Health Records," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 13, no. 6, p. e1549, 2021.
4. Hilal, W., Gadsden, S.A. and Yawney, J., "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," Expert Systems with Applications, vol. 193, p. 116429, 2022.
5. Amin, R. et al., "Healthcare Techniques Through Deep Learning: Issues, Challenges, and Opportunities," IEEE Access, vol. 9, pp. 98523-98541, 2021.
6. Nazir, S. et al., "A Comprehensive Analysis of Healthcare Big Data Management, Analytics, and Scientific Programming," IEEE Access, vol. 8, pp. 95714-95733, 2020.
7. Qiao, Y., Wu, K. and Jin, P., "Efficient Anomaly Detection for High-Dimensional Sensing Data with One-Class Support Vector Machine," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 404-417, 2021.
8. Choi, K. et al., "Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines," IEEE Access, vol. 9, pp. 120043-120065, 2021.
9. Yaacoub, J.P.A. et al., "Securing Internet of Medical Things Systems: Limitations, Issues, and Recommendations," Future Generation Computer Systems, vol. 105, pp. 581-606, 2020.
10. Suroor, N. and Misra, T., "Medical Insurance Fraud Detection," in Deep Learning in Internet of Things for Next Generation Healthcare, Chapman and Hall/CRC, pp. 182-193.
11. Alloghani, M. et al., "Prospects of Machine and Deep Learning in Analysis of Vital Signs for the Improvement of Healthcare Services," in Nature-Inspired Computation in Data Mining and Machine Learning, pp. 113-136, 2020.
12. Corny, J. et al., "A Machine Learning-Based Clinical Decision Support System to Identify Prescriptions with a High Risk of Medication Error," Journal of the American Medical Informatics Association, vol. 27, no. 11, pp. 1688-1694, 2020.
13. Javaid, M. et al., "Leveraging Lean 4.0 Technologies in Healthcare: An Exploration of Its Applications," Advances in Biomarker Sciences and Technology, vol. 6, pp. 138-151, 2024.
14. Elouataoui, W., El Mendili, S. and Gahi, Y., "An Automated Big Data Quality Anomaly Correction Framework Using Predictive Analysis," Data, vol. 8, no. 12, p. 182, 2023.
15. Erhan, L. et al., "Smart Anomaly Detection in Sensor Systems: A Multi-Perspective Review," Information Fusion, vol. 67, pp. 64-79, 2021.
16. Sunny, J.S. et al., "Anomaly Detection Framework for Wearables Data: A Perspective Review on Data Concepts, Data Analysis Algorithms and Prospects," Sensors, vol. 22, no. 3, p. 756, 2022.