FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, 2024, pp 283 - 296
DOI: https://doi.org/10.33003/fjs-2024-0803-2356

APPLICATION OF ALGORITHMS FOR ANOMALY DETECTION IN HEALTH-ENABLED SENSOR-CLOUD INFRASTRUCTURE

[1]Adigwe, A. R., *[1]Edje, E. A., [1]Omede, G., [1]Atonuje, O. E., [1]Akazue, M. I. and [2]Apanapudor, J. S.

[1]Department of Computer Science, Faculty of Science, Delta State University, Abraka Delta State, Nigeria
[2]Department of Mathematics, Faculty of Science, Delta State University, Abraka Delta State, Nigeria

*Corresponding authors' email: edjeabel@delsu.edu.ng

ABSTRACT

Real-time patient monitoring and early disease diagnosis are two ways that the healthcare industry is benefiting from the integration of sensors and cloud technology. In order to detect changes in patient's health, a variety of non-invasive sensors are applied to the skin to monitor various physiological parameters. The collected data are then wirelessly communicated to the cloud data center. However, the transmitted data are susceptible to several sources of interference called anomalies. Anomalies is when a sudden change occurs from the expected sensor data generated. This may be as a result of sensor faults, measurement faults, injection and alteration by malicious attackers. Therefore, this research tends to conduct a survey on existing algorithms or techniques used for the detection of anomalies in health-enabled sensor-cloud infrastructure. The processes adopted by the algorithms were identified and discussed exhaustively. In addition, the simulation setup and programming languages adopted to implement and evaluate the existing algorithms, followed by the limitations of the algorithms, which may lead to future research directions are captured in this paper. The outcome of the research shows that machine learning algorithms were predominantly adopted for detecting anomalies with the support of clustering and classification processes. Furthermore, Visual Basic.Net simulation tool and Python programming language was mostly adopted for experimentation and evaluation of the existing techniques. Limitations such as overfitting, under-fitting, computation complexity (time and memory space), and missing data are hindering the optimal performance of existing algorithm, which needs to be addressed in future researches.

Keywords: Cloud, Health Systems, Prediction Techniques, Deep/Machine Learning, Sensors

INTRODUCTION

Electronic sensors and sensor systems have become main producers of data, presently attaining yearly rates on the zettabyte scale (Cauteruccio *et al.,* 2021). Big data has emerged as a result of this ever-growing volume of data, which is notable not just for its bulk but also for its dynamic, velocity and heterogeneity (Edje and Latiff, 2021). A sensor is an apparatus that receives input from the external world and processes it. Light, heat, motion, moisture, pressure, and a variety of other environmental phenomena can all be used as inputs. Typically, the output is a signal that is either electronically communicated across a network for reading or additional processing, or it is transformed to a human-readable display at the sensor location (Edje *et al., 2021*). An essential components of the internet of things (IoT) are sensors. Through them, an ecosystem for gathering and analyzing data on a particular environment may be established, facilitating easier and more effective monitoring, management, and control (Erhan *et al,* 2020). Sensors are utilized in a variety of situations, including homes, fields, cars, aircraft, and factories. The physical and logical worlds are connected by sensors, which serve as the eyes and ears of a computing infrastructure that gathers, processes, and interprets sensor data.

The growth of using sensors in healthcare have increased tremendously over the last two decades. They are mostly used for retrieving health status information from patients. Medical sensors are either embedded in the human skin or are wearable on the physical body (Edje and Muhammad, 2021). Most medical sensors presently used in hospitals and care homes are wearable, which is more comfortable and acceptable by patients. There are variety of medical sensors that can measure a wide range of physiological attributes, which include heart rate (HR), body temperature (T◦), pulse, blood pressure (BP), respiration rate (RR), galvanic skin ratio (GSR), electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), and so on (Apanapudor ***et al.,*** 2023). These wearable sensors for monitoring vital signs have significant effects on public health.

The health status information retrieved from patients via medical sensors are usually transferred to the cloud-assisted computing for further processing (such as detection of anomalies) and analysis. This is because sensors are constrained with limited storage space and computational power to process the generated data. One possible solution that can address these challenges is cloud computing. Cloud computing uses the internet to allow rapid deployment, more flexible resource allocation, and economies of scale by delivering computer services such as servers, databases, networking, software, and data analytics (Apanapudor ***et al.,*** 2020). The goal of anomaly detection is to find data patterns that often different from expected behavior (Sanchez-Martrin *et al.,* 2019). This is important to know when looking for crucial details about how the system operates and identifying anomalies that are frequently uncommon or challenging to model or anticipate. Finding anomalies early on is essential to solving a variety of underlying issues that, if left unchecked, could have expensive repercussions (Aderibigbe, 2014). Examples includes predicting the onset of cancer, diabetes, heart beat rates, high blood pressure and low blood pressure.

Several research literature surveys have investigated the dynamics in healthcare sensor-cloud infrastructure over the last eight years. For instance, Dang et al., (2019) conducted a survey on emerging IoT and cloud components, applications and market trends in the healthcare sector. Also, considering the use of wearable devices applied in the healthcare sector, and the discovery of e-health policies that supports a viable development of IoT and cloud infrastructure in the healthcare sector. Aceto *et al.,* (2020)., surveys the application of IoT and cloud technologies in the health 4.0 industry. It further analyzes their benefits and novel cross-disciplinary challenges as well as lessons learned. Sofi *et al.,* (2022) investigated the use of wireless smart sensor network-enabled cloud for structural health monitoring. It analyzes the advances in data collection, manipulation, diagnostic, storage and retrieval phases of health monitoring for both commercial and academic purposes. The survey also indicates the lag in real-time processing of health monitoring devices notwithstanding improvement in academia.

Conversely, Fahd *et al.,* (2023) conducted a survey on cloud sensor architecture, applications and challenges in the health industry. It analyzes the integration of mobile computing, cloud computing and IoT sensor devices which is called Cloud of Things (CoT). Also, indicates the characteristic architecture, challenges with prospective solutions and applications of CoT in the health industry. Islam et al., (2023) conducted a review on cloud-IoT sensor based green healthcare system. analyzes the cutting-edge technology for interactive user interface, ensuring systems performance improvement ratio. It also indicates various types of wearable sensors used by patients to send their data and doctors can obtain those data in real-time. Hassani *et al.,* (2023) conducted a systematic review on sensor-cloud technologies for non-destructive and testing health monitoring. The research presented variety of sensing techniques which includes fiber optics, laser vibrometry, ultrasonic, thermography drones and next generation technologies. It also analyzes the limitations, advantages and disadvantages of the mentioned sensing technologies.

The previous reaches significantly analyzed and show the implications and prospects of applying sensor-cloud computing in health industry but are yet to investigates techniques and algorithms used for detecting data anomalies on health sensor-cloud computing. Therefore, this research tends to investigates the deployment of techniques and algorithms for detecting anomalies on health data in sensor-cloud. The contributions of this research are highlighted as follows.

i. Analysis of algorithms used for detecting anomalies on health enabled sensor-cloud computing.
ii. Processes adopted by the existing algorithms for detecting anomalies in health sensor-cloud computing.
iii. The simulation environment and programming languages adopted to implement and evaluated the performance of the exiting algorithms used for anomaly detection in health sensor-cloud.
iv. Challenges or weaknesses of the existing algorithms in active state, that can lead to feature research direction is also put forward.

The rest of this paper includes research methodology adopted to achieve the research contributions, research results which presents the analysis of the algorithms deployed, processes adopted by the existing algorithms, simulation environment and programming languages adopted for developing and evaluating the performance of the algorithms as well as their challenges. Also, highlighting the benefits and limitations of this research in the discussion section and ends with a conclusion.

MATERIALS AND METHODS

The research methodology used in existing literature reviews, authored by Dhanvijay et al., (2019) and Tahir *et al.,* (2020) was adopted to conduct the current study. Research methodology seeks to aid in comprehending the overall impact of utilizing several algorithms for the prediction of anomalies in the field of health sensor-cloud through an in-depth examination of these objectives.

The following are the primary Research Questions (RQ) that were formulated in order to investigate and conduct the current study.

i. What are the characteristics and functionalities of the algorithms in the health sensor-cloud?
ii. What opportunities and challenges of existing algorithms while detecting anomalies in health sensor-cloud?
iii. What processes or procedures adopted by algorithms for detecting anomalies in health sensor-cloud?
iv. What of simulation environment and programming languages adopted for development and performance evaluation of the algorithms?
v. What studies about algorithms for detecting anomalies in health sensor-cloud have been published?

Five major electronic research repositories (IEEE Xplore, Elsevier, Springer, Wiley Online Library, and Science direct) were searched for relevant articles that will aid the study. Nonetheless, this study also includes a few publications from MDPI and Hindawi that are somewhat relevant to the study. In order to conduct an automated in-depth text search by search engine screening and manual screening, we defined the following keywords for the searching process: "anomalies detection," "health sensor-cloud," and "anomalies detection Health sensor-cloud" in the context of the research domain. Using the table of contents for these keywords, the study was conducted using Boolean operators with the predefined keywords and within the scope of the formulated research questions to classify relevant articles. Also, the article in this paper was screened, filtered, and arranged according to the following inclusion criteria:

i. The paper's applicability to the use of algorithms for anomaly detection in health sensor cloud
ii. The articles written between 2016 and 2023
iii. should be written in English, as the primary language
iv. choosing only primary studies from relevant research
v. The paper should offer solid understanding based on the formulated research questions

In accordance with the inclusion criteria, a time restriction was setup for the search so that all relevant papers would be found and collected with an emphasis on the predetermined keywords. The paragraphs of the identified research papers were further scanned using the keywords in order to minimize their size and make them more manageable.

An estimate of 159 articles were gathered for the years 2016–2023 during the initial phase. In the next stage, a total of 90 articles were filtered out using screening based on keywords and titles. Using the predefined searching terms and the Boolean AND operator, the remaining articles were filtered based on the abstract in the last stage. The authors completed and chose the final 25 papers based on the inclusion criteria for additional research and analysis, taking into account each of the research questions that had been specified. Thanks to the full screening process from the first to the last phase as denoted in figure 1.

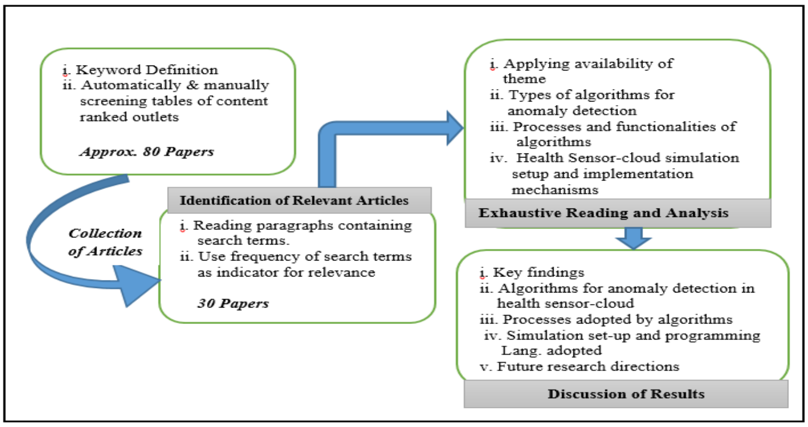

Figure 1: Research Methodology Structure (Tahir *et al.,* 2020)

The outcome of the research questions (RQ) as identified above, will be evaluated with the support of a table and statistical chart tools. The table will capture and highlights the various existing techniques adopted, processes, simulation/programming tools used for implementation and experimentation of the techniques and their limitations. Statistical chart tools will use to graphically show and comparing results based on performance.

RESULTS AND DISCUSSION

This section discusses and analyses the existing algorithms deployed to detects anomalies in the health sensor-cloud infrastructure. Followed by a detailed evaluation of results with the support of table, and other related scientific graphical toots (statistical chart tools).

Analyses of existing algorithms for Anomaly detection in Health Sensor-Cloud

Hybrid scheme to identify outliers in smart healthcare sensor cloud infrastructure was developed (Dwivedi et al., 2021). It consists of a Density-based Outlier and Gaussian Distribution techniques for the clustering and classification of sensed dataset to retrieve possible outliers. At the initial stage, the sensed dataset was pre-processed by removing excessive noise and duplicate data. The pre-processed dataset was clustered using the Density-based Scheme (DBS) by selecting irrelevant data point from the dataset and verify whether the points is within the low-density region or not. If the data point is within the low-density region, it is indicated as a possible outlier, otherwise it's a normal data. The identified possible data point outliers are further classified using Gaussian Distribution Approach (GDA) to retrieve the actual outliers.

A Context-aware Decision Approach was developed with the support of Adaptive Window-based (AWB) and Fuzzy-base Rule (FBR) techniques, for the monitoring of patient's activity in real time (Jayalakshmi and Gomath, 2020). The AWB technique is used to extract features from the sensed dataset be grouping the continuous data points into windows of static duration at a specific time length. This is affected by one half of the window length which converted to a single data point created for every window. Moreso, many data points are created from any particular data point. The extracted features retrieved are further classified using FBR to determine the status of patient health status based on static or dynamic action. Consequently, Su et al., (2019), implemented a framework that detects the abnormality in patient health statues. Sensed data collected from wearable senor device (s) are dispatched to a local server, where they are normalized to remove irregularities and classified using Linear Discriminant (LD) approach, based on pattern judgement to ascertain the current health status of the patient. An anomaly detection system (ADS) is developed for monitoring the health status of patients and environment (Said *et al.*, 2021). It uses the ordinary behavior with the support of a baseline to detect abnormality in patient's health. Thus, abnormality is detected if there is any possible deviation from the baseline. The abnormality data points are further classified using a Nonlinear Support Vector Machine algorithm. The SVM utilizes a hyperplane to segregates data into various classes (labels) and reformulate these data points using a statistical function called Kernel. Then, it defines an optimal borderline between the labels to obtain the actual abnormality data points. Consequently, Zhang et al., (2022) proposed a Time-efficient and privacy-aware anomaly detection framework for monitoring health status of students. Firstly, the generated health dataset of students is transformed into lightweight health indexes and stored in the cloud. Then, the similarity between each pair of indexes based on the health status of the students is computed. Indexes are further clustered based on their similarity level to obtain possible anomalies. Identical indexes are clustered into groups, while indexes that are not identical are regarded as anomaly. Signal analytic technique is proposed to detect anomalies in health status of patients (Nawaz et al., 2020). Patients dataset generated from Electrocardiogram (ECG) is filtered using Wavelet Adaptive (WA) Filtering approach to remove baseline movements for distortion reduction. Then, the filtered output data points are reconstructed using the signal data points and the baseline Coefficient Transform (CF) to retrieve the anomaly data points. Additionally, an intelligent system framework was developed to detect anomalies in the health status of patients (Nawaz and Ahmed, 2022). The generated datasets from electrocardiogram (ECG) is normalized or pre-processed using Wavelength Transform (WT) technique to remove noise and stored in the cloud. Then, an Auto-encoder (AE) algorithm is applied to discover possible anomalies. It compresses the pre-processed dataset into hidden space and decompresses the encoded dataset into the output layer. Thereafter, the decompressed dataset is compared with the input dataset to determine their propagation difference error to obtain possible anomaly data points. Hence, the retrieved possible anomaly data points are classified utilizing the Long Short-term Memory (LSTM) approach to obtain the actual anomalies in the health status of patients.

A cloud based remote monitoring system framework is designed for the detection of abnormalities in the health statues of patients (Shravan et al., 2017). It utilizes the Naïve Bayesian (NB) algorithm to determine the possibility of normal and abnormal data points in a given dataset. The NB deploys probabilistic approach by allocating labels from a given set of data points as either normal or abnormal. Then the Modified Weighted Average Method (MWAM) computed the decision parameter that monitors the degree of abnormality in the patient health status. A threshold value is defined by the decision parameter. If the value of the controlled variable is more than the threshold value then the abnormality is high, otherwise it is regarded as either medium or low.

A Graph Convolutional Network (GCN) technique is proposed for the detection of anomalies in smart healthcare (Wang et al., 2020). It detects malicious wearable sensor nodes that generates compromised or error datasets. It uses the vertices to represent sensor nodes and each edge signifies data sharing or transaction process. A probability prediction is inserted in the H_i matrix, which is responsible for monitoring any malicious nodes. Therefore, the data features retrieved from each node are transferred into the *Hi* matrix in conjunction with the adjacency matrix A. the anomaly features are detected by the probability prediction value.

A healthcare framework is developed for the detection and classification of physiological signals (Nawaz and Ahmed, 2021). It uses the Wavelet Scattering Transform (WST) technique for the classification of data generated by on-body noninvasive biomedical sensors. The classified data are test-trained with a deep leaning Auto-Encoder (AE) technique to obtain actual. Consequently, Althebyan et al., (2016) proposed a cloud-based support system for healthcare anomaly detection. It utilizes the Decision Support System (DSS) to detect abnormal data from the dataset generated by the wearable sensors. The abnormal data are forwarded to healthcare personnel for crucial healthcare intervention.

Healthcare framework is proposed for the detection of anomalies using big data techniques (Kaya et al., 2022). It utilizes four different types of classifier algorithms namely Random Cut Forest (RCF), Logistic Regression (LR), Neural Network (NN) and Naïve Bayes (NB) to detect abnormal data records from the entire dataset. Then, compares the performance outcome of the algorithms based on accuracy and execution metric. Performance evaluation results shows that the LR algorithm outperformed the other three algorithms bases on execution time. Conversely, Selvaraj et al., (2019) developed an Analogous Particle Swarm Optimization (APSO) approach for accurate selection virtual machines (VMs), to process the detection of anomalies in healthcare. The APSO obtains optimal VMs from pool of VMs by estimating and comparing the fitness function of all the VMs residing in the resource pool. Then, it updates the movement and position of each VM to determine the VM with the highest fitness value. The VMs with highest fitness values are used to predict abnormal data (Chronic Kidney Disease) from the dataset, using Neural network (NN) classifier algorithm.

A unified framework system for abnormality prediction for smart home is proposed (Grewal et al., 2019). The framework retrieves user behavior pattern data from electricity meter stationed at home. Firstly, it uses the X-means clustering approach to cluster similar patterns together without previous knowledge about the number of clusters. Then it uses the Local Outlier Factor (LOF) algorithm to determine the deviation of a specific data record from neighbors. After which, a prediction value is applied to detect anomaly. If the data point or record is higher than the prediction value (1.6), then it is classified as abnormal data point, otherwise it is regarded as normal data point.

A cloud-centric Internet of things (IoT) framework is developed for monitoring health conditions of students to detect abnormality (Verma et al., 2018). It detects the student's diseases (abnormality) by extracting the values or data generated from IoT sensors and medical record history. The data is further processed and validated using K-cross validation approach after which it is classified as normal or abnormal using Decision and K-nearest neighbor approaches. Conversely, Verma and Sood, (2018), further improved the cloud-centric IoT framework for the prediction of abnormality in sensed big data generation. The sensed dataset is segmented into clusters, using the Cross-modality Search (CMS) and Regression Methods to identify several severity attributes. Then, the severe attributes which are also regarded as possible anomaly are classified to obtain actual abnormal data using various classification algorithms.

A Markov Chain (MC) approach is proposed for the monitoring patient's health condition in other to detect abnormality (Salem et al., 2021). The MC technique is formulated from the Root Mean Square Error (RMSE) between the measured values for entire features. The generated features are converted into a univariate times series, which is categorized into overlapping sliding windows by the RMSE. In each sliding window, the mutual probability of the categorized RMSE values is computed to determine if a change has occurred or not. Therefore, the number of deviated features is utilized to differentiate faulty values or event occurrence from the whole measured features when an imminent change is detected over K successive windows. The faulty values or events detected are classified as abnormal while the other values are normal.

A monitoring system is proposed to detect abnormality in the health condition of patients (Hridhya et al., 2019). Sensed data generated from sensors are transferred to a controller where they are preprocessed to remove excessive noise. Thereafter, the improved sensed dataset is trained test/trained using Random Forest algorithm to obtain abnormal data records. Consequently, a detecting illegal behavior (DIB) system is proposed for safe guarding medical Control services against external attacks (Fang et al., 2020). Packets generated are converted into supervised datasets and a window with time length of W is set to segregate the sequence packets. For each widow a feature vector is extracted which is made up of five variables. To determine whether an event has occurred, a feature vector is labeled as 1 while 0 lebel is assigned to no-event.

Anomaly prediction based on medical sensor data stream framework is developed to improve the quality of service rendered to patients (Sun et al., 2021). Sensed dataset retrieved from the ECG devices are transmitted to the cloud, where they are processed. The deep learning based Stack Long-Short Term Memory Architecture Prediction (SLAP) technique is deployed in the cloud to train/test the sensed dataset. SLAP consists of two sub-layers of the conventional LSTM algorithm, where each edge linking two cells has a weight. The hidden layer of LSTM connects the other hidden layer, so that the upper layer outputs data records is collected to the lower layer. Then, a Multivariate Gaussian Distribution (MGD) algorithm is deployed to differentiate normal from abnormal data samples, by using a threshold method. Consequently, Yatbaz et al. (2021), proposed a Lightweight Convolutional Neural Network (LCNN) model for the prediction of anomalies in patient health condition. The generated sensor dataset is initially normalized using the Min-Max (MM) technique and de-noised using Baseline Wander Removal (BWR) technique. Then, the de-noised data features are labeled for training purpose with the utilization LCNN model, which output the end result as either abnormal or normal as denoted.

A Bayesian Network (BN) Classifier and Thread Protocol (TP) algorithms is developed for the prediction of anomalies in patient heath condition (Tanwar et al., 2020). The tread protocol is used to transmit sensed data retrieved from Wearable Area-sensor Network (WABN) to the cloud data center. Therefore, the sensed dataset is test/trained using the BN algorithm to predict the actual abnormal data from the normal data records. Conversely, Arpaia et al. (2022) proposed a Deep Learning-based Long-Short-Term-Memory (LSTM) Auto-encoder for the detection of anomalies in patients' health conditions. Initially, a Multivariate Linear Regression (MLR) algorithm is utilized to evaluate the values of the dataset in the cloud database, then the LSTM Auto-encoder algorithm is used to process the entire dataset. The LSTM uses thresholds, which were regulated by the reconstruction errors that the Auto-encoder processes in the training level. If the data record exceeds the threshold value then it is regarded as anomalies, otherwise is a normal data record.

Results

The results obtained from conducting this research survey are analyzed and discussed extensively as shown in table 1 (Appendix section). These results were further evaluated using statistical tools for better representations and more insight. The table highlights the types of algorithms or techniques adopted, processes adopted, challenges resolved, outcome, simulation packages, benchmark, Metrics and limitations of the said existing techniques.

The type of Algorithm and Processes Adopted by existing Algorithms

The results of the literature review as presented in table 1, indicate that machine learning algorithms were predominantly deployed on an average of 70%, for anomaly detection in healthcare sensor-cloud as compared to other types of techniques, denoted in figure 2a. On the other hand, clustering/classification processes outperforms other types of processes adopted by the existing techniques, as captured in figure 2b.

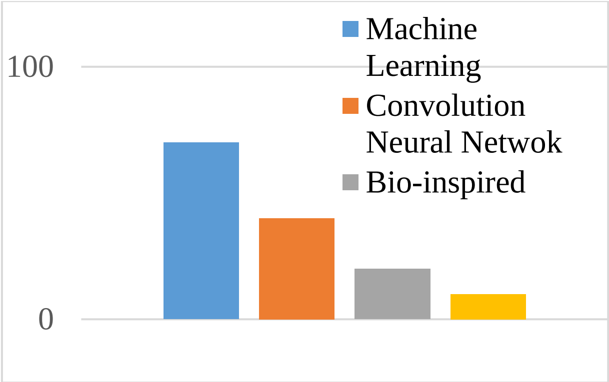

Figure 2a: Comparing Categories of Techniques

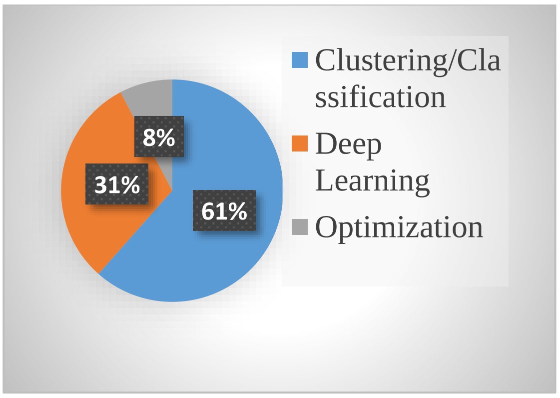

Figure 2b: Comparing of process Outcomes

Clustering process is mostly used for the selection and extraction of pertinent characteristics from the dataset, the classification process is primarily utilized for the identification of anomalies in a given dataset.

Programming languages and Simulation Tools adopted

It was also discovered that Python and Visual Basic.Net tools were mostly adopted for implementation and conduct experimentation on the techniques as illustrated in figure 3a and 3b.

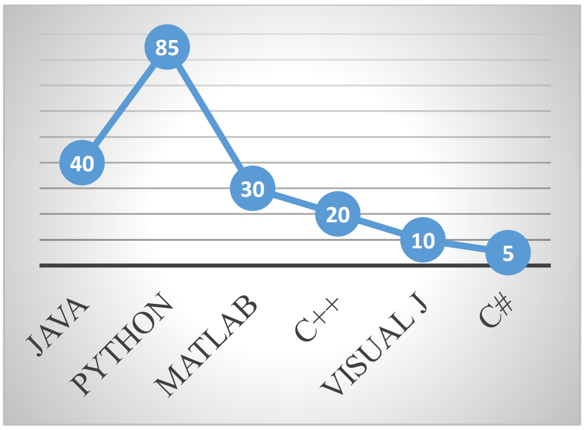

Figure 3a: Programming Lang. Usage Comparison

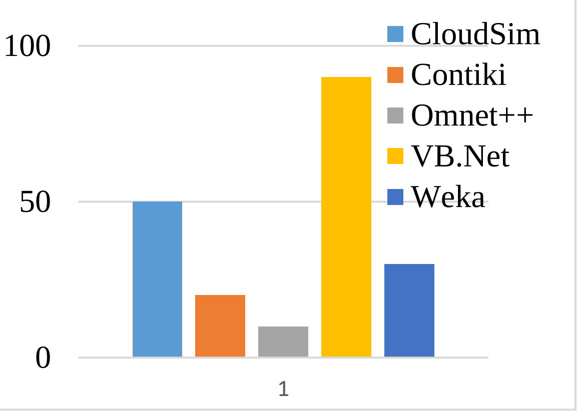

Figure 3b: Simulation Tools Usage Comparison

Python is a programming language that may be used to make websites, automate processes, analyze and visualize data, and create software. Visual Basic.NET depends on the.NET framework, applications created in the language have high levels of scalability and dependability.

Limitation of the Existing Techniques for Future Researches

Despite the performance improvement of the existing models as denoted in table 1, they are still prone to some challenges that could lead to future research direction. These challenges include data over-fitting and under-fitting, missing data issues and computational complexity (memory usage and untimely execution). Overfitting/under-fitting were discovered as the most prevailing challenges of the existing models than other related challenges, with on average of estimate of 80%, as denoted in figure 4.

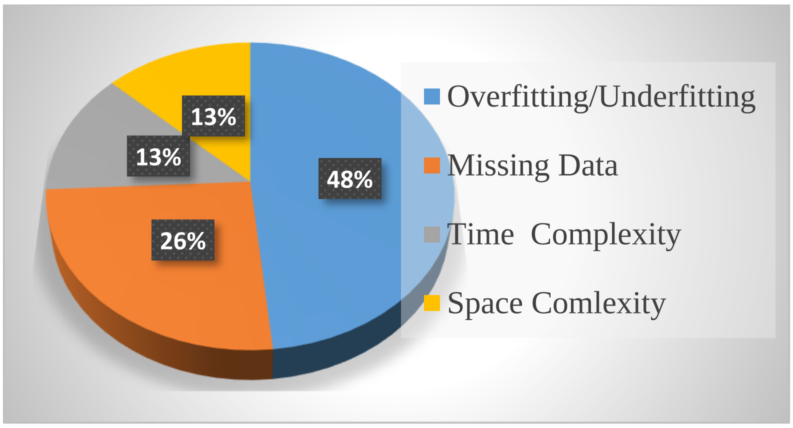

Figure 4: Limitations of Existing Techniques Comparison

When our machine learning model attempts to cover all the data points in the dataset or more than is necessary, it is said to be overfitting. As a result, the model begins to cache noise and erroneous values found in the dataset, all of which lower the model's accuracy and efficiency.

When our machine learning model is unable to identify the underlying trend in the data, under fitting happens. The training data feed can be terminated early to prevent the model from overfitting, which could result in the model learning insufficiently from the training data. It might therefore be unable to determine how well the data fits the dominating pattern. A problem with missing data is that it causes uncertainty in your analysis. A data set contains missing data (also referred to as missing values) when some of the observations are blank. Moreover, an observation is considered unusual if it has missing data for a variable. As a result, any study that makes the assumption that the missing value neatly fits into the remaining data is flawed.

Determining the computational resources needed to solve a problem involving time, space, or communication as well as the potential and constraints of algorithmic efficiency are the two main

CONCLUSION

Generally, the research show that machine learning algorithms are majorly deployed to detect anomalies in health sensor cloud infrastructure. Anomalies as it applies in health, is the prediction of impending diseases in life of an individual or prospective patient. Most of the algorithms deployed for this purpose yielded significant performance results based accuracy, precision and recall, specificity with limited execution time. this means that machine learning algorithms are reliable and efficient for processing health related data irrespective of the amount (Big data) of datasets in question. In this study, a thorough literature analysis of various algorithms used for the detection of anomalies in health enabled sensor-cloud was presented. It was observed that Visual Baic.Net simulation tool and Python programming language were mostly adopted for the implementation and experimentation of the existing techniques. Machine learning models happens to be the most sort after techniques, while under-fitting and over-fitting are still the most prevailing challenges of existing techniques adopted for anomaly detection in health sensor-cloud paradigm. Also, clustering and classification processes were mostly adopted by the existing techniques to resolve pending challenges for the detection of anomalies in the health sensor-cloud. The limitations of the existing algorithms for anomaly detection in the health sensor-cloud setting may be resolved by developing future framework models and techniques, and this study contributes to identifying the research challenges for that development. An effort in the future to develop some state-of-the-art techniques for the detection of anomalies in health sensor-cloud infrastructure, specifically monitoring the health status of elderly patients in real time basis.

REFERENCES

Abel, Edje Efetobor and Muhammad, Abd Latiff Shafie (2021). Management of WSN-enabled Cloud Internet of Things: A Review, International Journal of Computing and Digital Systems, 10(1), Pages 354-372Adnan Tahir, Fei Chen, Habib Ullah Khan, Zhong Ming Apanapudor, J. S., Aderibigbe,

F. M. and Okwonu, F. Z. (2020). An Optimal Penalty Constant for Discrete optimal control Regulator Problems, *Journal of Physics: Conference Series*, 1529(4), Pages 042-073.

Apanapudor, J. S., Umukoro, J., Okwonu, F. Z. and Okposo, N. (2023). Optimal Solution Techniques for Control Problem of Evolution Equations, *Science World Journal*, 18(3). Pagges 503-508.

Arshad Ahmad, Shah Nazir and Muhammad Shafiq (2020). A Systematic Review on Cloud Storage Mechanisms Concerning e-Helathcare Systems, *Sensors (MDPI)*, 20, 1-32.

Althebyan, Q., Yaseen, Q., Jararweh, Y., & Al-Ayyoub, M. (2016). Cloud support for large scale e-healthcare systems. *Annales Des Telecommunications/Annals of Telecommunications*, *71*(9–10). https://doi.org/10.1007/s12243-016-0496-9

Arpaia, P., Crauso, F., de Benedetto, E., Duraccio, L., Improta, G., & Serino, F. (2022). Soft Transducer for

Patient's Vitals Telemonitoring with Deep Learning-Based Personalized Anomaly Detection. *Sensors*, *22*(2). https://doi.org/10.3390/s22020536

Cauteruccio, F., Cinelli, L., Corradini, E., Terracina, G., Ursino, D., Virgili, L., Savaglio, C., Liotta, A., & Fortino, G. (2021). A framework for anomaly detection and classification in Multiple IoT scenarios. *Future Generation Computer Systems*, *114*. https://doi.org/10.1016/j.future.2020.08.010

Dwivedi, R. K., Kumar, R., & Buyya, R. (2021). A Novel Machine Learning-Based Approach for Outlier Detection in Smart Healthcare Sensor Clouds. *International Journal of Healthcare Information Systems and Informatics*, *16*(4). https://doi.org/10.4018/IJHISI.20211001.oa26

Edje E. Abel and Muammad Shafie Abd Latiff (2021). The utilization of algorithms for cloud internet of things application domains: a review, Frontiers of Computer Science (Springer), 15(3), Pages 1-27

Edje E. Abel, Abd Latiff Muhammad Shafie and Weng Howe Chan (2021). Deployment of internet of things-based cloudlet-cloud for surveillance operations, IAES International Journal of Artificial Intelligence (IJ-AI), 10(1), Pages 24-34

Erhan, L., Ndubuaku, M., di Mauro, M., Song, W., Chen, M., Fortino, G., Bagdasar, O., & Liotta, A. (2021). Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion*, *67*, 64–79. https://doi.org/10.1016/J.INFFUS.2020.10.001

Fahd Alhaidari, Atta Rahman and Rachid Zagrouba (2023). Cloud of Things: Architecture, Applications and Challenges, *Journal of Ambient Intelligence and Humanized Computing (Springer)*, 14(2023), Pages 5957-5975.

Fang, L., Li, Y., Liu, Z., Yin, C., Li, M., & Cao, Z. J. (2021). A Practical Model Based on Anomaly Detection for Protecting Medical IoT Control Services against External Attacks. *IEEE Transactions on Industrial Informatics*, *17*(6). https://doi.org/10.1109/TII.2020.3011444

FM Aderibigbe and JS Apanapudor (2014). On the Extended Conjugate Gradient Mehtod (ECGM) Algorithm for Discrete Optimal Control Problems and some of its features, *IOSR Journal of Mathematics*, 10(3). Pages 16-22.

Martins, P., Reis, A. B., Salvador, P., & Sargento, S. (2020). Physical Layer Anomaly Detection Mechanisms in IoT Networks. *Proceedings of IEEE/IFIP Network Operations and Management Symposium 2020: Management in the Age of Softwarization and Artificial Intelligence, NOMS 2020*. https://doi.org/10.1109/NOMS47738.2020.9110323

Grewal, A., Kaur, M., & Park, J. H. (2019). A unified framework for behaviour monitoring and abnormality detection for smart home. *Wireless Communications and Mobile Computing*, *2019*. https://doi.org/10.1155/2019/1734615

Guiseppe Aceto, Valerio Persico and Antonio Pescape (2020). Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for *Healthcare 4.0, Journal of Industrial Information Integration (Elsevier), 18, (2020), Pages 100-129.*

Hridhya A. P., Periasamy and Rahul I. R. (2019). Patient Monitoring and Abnormality Detection Along with an Android Application, *International Journal of Computer Communication and Informatics, 1(1), Pages 52-57*

Jayalakshmi, M., & Gomathi, V. (2020). Pervasive health monitoring through video-based activity information integrated with sensor-cloud oriented context-aware decision support system. *Multimedia Tools and Applications*, *79*(5–6). https://doi.org/10.1007/s11042-018-6716-8

Minh Dang, L., Piran, M. J., Han, D., Min, K., & Moon, H. (2019). A survey on internet of things and cloud computing for healthcare. *Electronics (Switzerland)*, *8*(7). https://doi.org/10.3390/electronics8070768

Motaharul Islam and Zaheed Ahmed Bhuiyan (2023). An Integrated Scalable Framework for Cloud and IoT Based Green Healthcare System, Access (IEEE), 11(2023), Pages 22266-22282.

Mrinai M. Dhanvijay, Shailaja C. Patil (2019). Internet of Things: A survey of enabling technologies in healthcare and its applications, *Computer Networks (Elsevier)*, 153, 113-131

Nawaz, M., & Ahmed, J. (2022). Cloud-based healthcare framework for real-time anomaly detection and classification of 1-D ECG signals. *PLoS ONE*, *17*(12 December). https://doi.org/10.1371/journal.pone.0279305

Nawaz, M., Ahmed, J., Abbas, G., & Ur Rehman, M. (2020). Signal Analysis and Anomaly Detection of IoT-Based Healthcare Framework. *2020 Global Conference on Wireless and Optical Technologies, GCWOT 2020*. https://doi.org/10.1109/GCWOT49901.2020.9391621

Osman Salem, Khalid Alsubhi, Ahmed Mehaoua and Raouf Boutaba (2021). Markov Models for Anomaly Detection in wireless Body Area Networks for Secure Health Monitoring, *IEEE Journal of Selected Areas in Communications*, 39(2), Pages 526-540

Prabal Verma, Sandeep K. Sood and Sheetal Kalra (2018). Cloud-centric IoT based student healthcare monitoring framework, *Journal of Ambient Intelligence and Humanized Computing (Springer)*, 9, Pages 1293-13009

Rahadian, Hassani., Bandong, S., Widyotriatmo, A., & Joelianto, E. (2023). Image encoding selection based on Pearson correlation coefficient for time series anomaly detection. *Alexandria Engineering Journal*, *82*, 304–322.

Said, A. M., Yahyaoui, A., & Abdellatif, T. (2021). Efficient anomaly detection for smart hospital iot systems. *Sensors (Switzerland)*, *21*(4). https://doi.org/10.3390/s21041026

Sánchez-Martín, J. M., Rengifo-Gallego, J. I., & Blas-Morato, R. (2019). Hot Spot Analysis versus Cluster and Outlier Analysis: An enquiry into the grouping of rural accommodation in Extremadura (Spain). *ISPRS International Journal of Geo-Information*, *8*(4). https://doi.org/10.3390/ijgi8040176

Selvaraj, A., Patan, R., Gandomi, A. H., Deverajan, G. G., & Pushparaj, M. (2019). Optimal virtual machine selection for anomaly detection using a swarm intelligence approach. *Applied Soft Computing Journal*, *84*. https://doi.org/10.1016/j.asoc.2019.105686

Sofi A., Jane J. Regita, Bhagyesh Rane and Hieng Ho Lau (2022). Structural health monitoring using wireless smart sensor network – An overview, *Mechanical Systems and Signal Processing (Elsevier)*, 163(2022), Pages 108-113.

Su, C. R., Hajiyev, J., Fu, C. J., Kao, K. C., Chang, C. H., & Chang, C. ter. (2019). A novel framework for a remote patient monitoring (RPM)system with abnormality detection. *Health Policy and Technology*, *8*(2). https://doi.org/10.1016/j.hlpt.2019.05.008

Sun, L., Yu, Q., Peng, D., Subramani, S., & Wang, X. (2021). Fogmed: A fog-based framework for disease prognosis based medical sensor data streams. *Computers, Materials and Continua*, *66*(1). https://doi.org/10.32604/cmc.2020.012515

Tanwar, S., Vora, J., Kaneriya, S., Tyagi, S., Kumar, N., Sharma, V., & You, I. (2020). Human Arthritis Analysis in Fog Computing Environment Using Bayesian Network Classifier and Thread Protocol. *IEEE Consumer Electronics Magazine*, *9*(1). https://doi.org/10.1109/MCE.2019.2941456

Verma, P., & Sood, S. K. (2018). Cloud-centric IoT based disease diagnosis healthcare framework. *Journal of Parallel and Distributed Computing*, *116*.

Xie, Y., Zhang, K., Kou, H., & Mokarram, M. J. (2022). Private anomaly detection of student health conditions based on wearable sensors in mobile cloud computing. *Journal of Cloud Computing*, *11*(1). https://doi.org/10.1186/s13677-022-00300-x

Yatbaz, H. Y., Ever, E., & Yazici, A. (2021). Activity Recognition and Anomaly Detection in E-Health Applications Using Color-Coded Representation and Lightweight CNN Architectures. *IEEE Sensors Journal*, *21*(13)

APPENDIX

Table 1: Results of the research highlighting solutions to key research questions (RQ)

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
Cloud-based healthcare framework for real-time anomaly detection and classification of 1-D ECG signals. Nawaz and Ahmed (2022)	Weave Transform (WT) algorithm, Auto-Encoder (AE) algorithm and Long Short-term Memory (LSTM) algorithm	Deep learning and Classification	Inability to detect anomalies in real time healthcare data stream	Improved the detection of anomalies in health status of patients based on accuracy, mean absolute error, precision and recall.	Python	K Nearest Neighbor (KNN), Support Vector Machine (SVM) and Ensemble Bagged Tree algorithms	Accuracy, Precision, Mean Absolute Error and Recall.	Challenges with temporal and morphological variations in various patients.
Signal Analysis and Anomaly Detection of IoT Based Healthcare Framework. Nawaz et al. (2020)	Wavelet Adaptive (WA) Filtering Approach and Baseline Coefficient Transform (BCF) algorithm	Signal Processing	Excessive Distortion in signal data points	Improved anomaly detection rate based on heart rate and human fatigue.	Virtual Instrument Software Architecture	Not Specified	Heart Beat Rate, Heart Rate Mean, Time Mean and Interval Mean	Not Specified
Abnormality Detection on Vital Parameters using Modified Weighted Average Method in Cloud. Shravan et al., (2017).	Naïve Bayesian Algorithm and Modified Weighted Average Method (MWAM).	Classification	Not considering patients history data records	Improved respiratory rate and timeliness in detecting high, medium and low abnormalities in patients' health status.	Open Nebula, Python	Simple K-means and Naïve Bayesian	Time efficiency and Respiratory Rate.	Not specified
GuardHealth: empowered secure data management and Graph Convolutional Network enabled anomaly detection in smart healthcare. Wang et al., 2020	Graph Convolution Neural Network (GCN) Technique	Deep learning	Compromised healthcare data integrity	Improved detection of anomaly sensor nodes based on precision and transaction epoch.	Python 3.9	Logistic Regression and Multilayer Perception algorithms.	Precision and Transaction Epoch.	Limited in accurate detection of anomalies in small amount of dataset.

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
Anomaly Detection of Physiological Signals in IoT-Based Healthcare Framework. (Menaa Nawaz and Jameel Ahmed, 2021)	Wavelet Scattering Transform and Auto-Encoder Techniques	Classification and Deep learning.	Inability of anomaly detection in large dataset	Improved detection rates based on accuracy and average absolute error loss	R-Studio	Naïve Bayes, Support Vector Machine	Accuracy and Average Absolute Error Loss	Missing data
Cloud support for large scale e-healthcare systems (Althebyan et al., 2016)	Decision Support System (DSS)	Classification	Inability of anomaly detection in large dataset	Improved detection rates based on conserving energy usage and time efficiency.	CloudExp Simulator	Not Specified	Execution time and Energy consumption rate	Under-fitting
Anomaly Detection and Performance Analysis by Using Big Data Filtering Techniques For Healthcare on IoT Edges (Şükrü Mustafa Kaya et al., 2022)	Randdom Cut Forest (RCF), Logistic Regression (LG), Neural Network (NN) and Naïve Bayes (NB) algorithms	Classification	Global optima search space	Improved detection rates based on accuracy and execution time.	Python		Accuracy, Racall, Precision and Execution Time	Challenges in classification process by NN algorithm
Optimal virtual machine selection for anomaly detection using a swarm intelligence approach (Selvaraj et al., 2019)	Analogous Particle Swam Optimization (APSO) and Neural Network Techniques	Optimization and Classification	Inability of optimal VM selection for anomaly detection operation	Improved optimal VM selection for anomaly prediction rates based on accuracy and Mean Squared Error (MSE) and reducing execution time (ET) and Resource utilization (RU).	CloudSim	PSO, Gnetic Algorithm (AG), K-nearest Neighbors (KNN), Decision Tree (DT), Fuzzy-c means (FC)	Execution time, Resource Utilization, Mean Squared Error, Accuracy.	Missing data

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
A Unified Framework for Behaviour Monitoring and Abnormality Detection for Smart Home (Grewal et al., 2019)	X-means and Local Outlier Factor algorithms	Clustering and Classification	Inability to detect abnormally due to over-fitting.	Improved detection rates based on energy conservation and Execution time.	Visual Basic.Net and Python	K-nearest Neighbor, Support Vector Machine and Naïve Bayes	Execution time, Energy conservation, Accuracy, Precision, Recall and FI-score	Inability to work on real-time basis.
Markov Models for Anomaly Detection in Wireless Body Area Networks for Secure Health Monitoring (Salem et al., 2021)	Markov Chain Approach	Classification	Vulnerability of data capturing by faulty sensors, injection and alteration by malicious attackers.	Improved anomaly detection rates based on execution time and raised alarm.	Python and Weka	K-nearest Neighbor, Support Vector Machine and J48	ROC, Root Mean Square Root (RMSE), True Positive Rate (TPR), False Alarm Rate (FAR	Computationally intensive based on time and memory usage.
Cloud-centric IoT based student healthcare monitoring Framework (Verma et al., 2018)	Decision Tress (DT) and K-nearest Neighbor (KNN)	Classification	Local minima search space	Improved abnormal detection rates based on response time, accuracy.	Weka, MySQL	K-nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT) and Naïve Bayes (NB)	Accuracy, Sensitivity, Specificity, Average Response Time.	Unable to adapt to big data generation.
A Practical Model Based on Anomaly Detection for Protecting Medical IoT Control Services Against External Attacks (Fang et al., 2020)	Fuzzy Core Vector Machine (FCVM) technique	Classification	Un-authorized access to medical information and devices.	Improved detection n rates based on average response time and accuracy.	Visual Basic.Net	Support Vector Machine (SVM) and Isolation Forest (IF)	Accuracy and Average Response Time	Global optima search space challenge.
Patient Monitoring and Abnormality Detection Along with an Android Application (Hridhya et al., 2019).	Random Forest Approach	Classification	Inability to manage health information of patients retrieved from sensors.	Improved monitoring and detection rates based on response time.	C# and C	Not Specified	Response Time	Deficiency in accurate detection of abnormality.

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
A Ubiquitous architecture for wheelchair fall anomaly detection using low-cost embedded sensors and isolation forest algorithm (Yoysuf and Kadri, 2023)	Isolation Forest (IF), Zero Angular Rate Update (ZART), Complementary Filter (CF) and Relief Algorithms.	Classification	Prevention of wheelchair fall	Improved detection rate based FI-score, AUC-ROC scores	JAVA	Local Outlier Factor (LOF), OC-Support Vector Machine and Convolution Neural Network (CNN).	G-mean, Area Under Curve and FI-score.	Over-fitting
Soft Transducer for Patient's Vitals Telemonitoring with Deep Learning-Based Personalized Anomaly Detection (Arpaia et al., 2022)	Multivariate Linear Regression (MLR) algorithm and Long-Short Term Memory (LSTM) Auto-encoder algorithm	Deep Learning	Challenges of overfitting and under fitting.	Improved anomaly detection rate based on accuracy and FI-score.	JAVA	Not Specified	Area Under Curve (AUC), FI-score and Binary Accuracy.	Not Specified
Human Arthritis Analysis in Fog Computing Environment Using Bayesian Network Classifier and Thread Protocol (Tanwar et al., 2020)	Bayesian Network algorithm and Thread Protocol.	Classification	Challenges of improper detection of anomaly in patient health condition.	Improved detection rates based on response to time and packet delivery time.	OMNet ++ Simulator and Python	Not Specified	Packet Delivery Ratio, Response Time and Packet Delivery Rate	Inability to predict anomaly in large dataset.
FogMed: A Fog-Based Framework for Disease Prognosis Based Medical Sensor Data Streams (Sun et al., 2021)	Stack Long-Short Term Memory Architecture Prediction (SLAP) and Multivariate Gaussian Distribution (MGD) algorithms	Deep learning and classification	Challenges of quality of service	Improved quality of service for anomaly detection based on response time and minimum usage of memory space.	Python	Recursive Neural Network (RNN) and Long-Short Term Memory (LAST) techniques	Response Time, FI-score, Accuracy and Recall.	Energy constrained devices.

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
Cloud-centric IoT based disease diagnosis healthcare framework (Verma and Sood, 2018)	Cross-modality Search, Regression and K Nearest Neighbor (KNN)	Classification	Challenges with global search space	Improved anomaly detection on global search space.	WEKA 3.0	Support Vector Machine (SVM), A-Neural Network (A-NN), K-nearest Neighbor (K-NN)	FI-score, Accuracy, Response Time, Specificity and Sensitivity	Computational complexity based on time constraint
Activity Recognition and Anomaly Detection in E-Health Applications Using Color-Coded Representation and Lightweight CNN Architectures (Yatbaz et al., 2021)	Colour-coding Convolution Neural Network (CC-CNN) model.	Deep Learning	Challenges with computational complexity	Improved anomaly detection base accuracy and minimum amount of memory usage and complexity.	Python		FI-score and Accuracy	Under-fitting
A Novel Machine Learning-Based Approach for Outlier Detection in Smart Healthcare Sensor Clouds. Dwivedi et al., (2021)	Density-based (DBA) Technique and Gaussian Distribution Approach (GDA)	Clustering and Classification	Error data points from faulty sensors and compromised data points integrity by intruders.	Improved outlier detection rate based on accuracy, recall, F1-score, throughput and efficiency.	Python	Mean Shift (MS) and Support Vector Machine (SVM)	Accuracy, Recall, FI-score, Throughput and Efficiency	Unable to detect outliers in global optimal search space.
Pervasive health monitoring through video-based activity information integrated with sensor-cloud oriented context-aware decision support system. Jayalakshmi and Gomathi (2020)	Adaptive Window-based Technique and Fuzzy Dynamic weighted Adaptation	Feature Selection and Classification	Blurred or excessive noise	Improved the detection of health status of patients based on accuracy, precision, specificity and detection rate.	Weka 3.0 and Java.	Standard CBR and Fuzzy Dynamic Weights Algorithm	accuracy, precision, specificity and detection rate.	Not Specified
A novel framework for a remote patient monitoring (RPM) system with abnormality detection. Su et al. (2019)	Remote Patient Monitoring (RPM) Framework and Linear Discriminant Technique	Classification	Local optima search space	Improved detection rates and efficiency based on timeliness.	Philip VM6 Physiology Device, ECG,	Not Specified	Detection Rates	Not Specified

Article Title	Algorithm	Processes	Challenges Resolved	Outcome	Simulation Package	Benchmark	Metrics	Weakness
Efficient Anomaly Detection for Smart Hospital IoT Systems. Said et al., (2021)	Anomaly Decision Manager (ADM) algorithm and Non-linear Support Vector Machine (SVM) Algorithm	Classification	Irregularities amidst captured dataset that leads to inaccurate detection of patient health status.	Improved abnormality even detection accuracy based on Increase Temperature Level and Heart Attack.	Contiki Cooja	Not Specified	Anomaly Detection Rate (ADR)	Inefficiency base on timeliness
Private anomaly detection of student health conditions based on wearable sensors in mobile cloud computing. Zheng et al., (2022)	Ano-Det Algorithm	Clustering	High data transmission cost and sensitivity of large dataset impedes the feasibility of anomaly	Improved accurate detection of anomalies in students health status time efficiency, and accuracy	MATLAB	Not Specified	Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)	High energy usage and inability to detect anomaly in Global optima space.

©2024 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.