An Evaluation of Outlier Detection Using Machine Learning in Medicine

Gaikwad Mahesh Parasharam [1] , Dr. Harsh Lohiya [2]

[1] Research Scholar, Dept. of Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India.

[2] Research Guide, Dept. of Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India.

Article Info
Page Number:952 - 965
Publication Issue:
Vol 70 No. 2 (2021)

Abstract

Detecting outliers is a serious issue that has been investigated in a number of academic and application fields. To effectively detect outliers, researchers are still developing reliable systems. In the area of medical research, the outlier detection problem has several relevant applications. Big data is rapidly being acknowledged and valued by individuals as a result of the extraordinarily quick increase of data in numerous industries. Numerous parties have taken an interest in medical big data, which can best illustrate the utility of big data. Numerous notes about a patient's ailments, therapies, and lab results can be found in their medical records. typically incorporate several different sorts of data and generate a lot of information. These databases can offer crucial data to enhance hospital management and clinical decision-making. Some specifics discovered in medical databases are rarely present in other non-medical sources. In this situation, anomalous patterns in health records (such as issues with data quality) can be found using outlier detection techniques, which will lead to better data and knowledge for decision-making.

In the field of data analysis, outlier detection has long been a key idea. The direct relationship between data outliers and real-world abnormalities, which are of significant interest to analysts, has recently been realised in a number of application domains. Finding patterns in data that deviate from predicted typical behaviour is known as outlier detection. The secondary sources are where the study's data was gathered. The primary goal of this study is to examine outlier detection in medical data using machine learning techniques.

Article History
Article Received*: 05 September 2021*
Revised*: 09 October 2021*
Accepted*: 22 November 2021*
Publication*: 26 December 2021*

Keywords: Outlier detection, Medical databases, Machine learning approaches, Disease diagnosis.

1. Introduction

A key idea in the study of medical data is outlier detection. The most potential uses of data mining are the complicated correlations that emerge between a patient's diabetic symptoms, diagnoses, and behaviour. Data objects that do not adhere to the overall behaviour of the data may be found in a data base. Outlier mining is the process of analysing outlier data, which includes these data objects. Finding new information from a big collection of data is the goal of data mining. Because there are many different types of sequences and there are many different ways to characterise outliers in sequences, there are many possible problem formulations for the problem of outlier detection for data mining. Outliers are typically ignored by data mining techniques as noise or exceptions(A.

Nisioti, 2021). One of the most crucial steps in data pre-processing is the treatment of outlier observations in a data set for two reasons. First, outlier observations can significantly affect an analysis's findings. Second, outliers are frequently the result of measurement or recording errors; some of them may even reflect interesting events or other important things from the perspective of the application domain.

Medical data analysis made it possible to diagnose patients by predicting clinical outcomes using biological data, such as biomarkers or disease patterns. Classification and prediction-based machine-learning algorithms are employed because medical data are challenging to analyse using traditional statistical methods due to their high dimensionality and sparseness. To decrease the illness features and improve the accuracy of the algorithms, a feature selection technique based on correlation was also applied. The nature of this dataset appears to have a significant impact on algorithm performance, yet there is no machine learning method that would consistently outperform one another. The correlation-based feature selection strategy, however, increased the prediction accuracy of all models by eliminating pointless genes. Several thousand to several tens of thousands of gene/protein sequences are represented in medical data. Each piece of medical data is scanned and converted into continuously normalised data(Alakus TB, 2020). The large dimensionality and unbalanced nature of the medical datasets are their key problems. Traditional machine learning classifiers use a small collection of features and have high error and true negative rates for disease prediction. Many thousand to several tens of thousands of gene/protein sequences are represented in a biomedical dataset. Every dispersed set of medical data is scanned and converted into normalised continuous data. The large dimensionality and unbalanced nature of the distributed medical data are the key problems. Traditional machine learning classifiers use a small collection of features and have high error and true negative rates for disease prediction. In order to handle the uncertain data effectively utilising the information gain, the impurity of the random samples is determined using the well-known information theory measure, entropy. The amount of contaminants in a group of samples is measured by entropy. As an evaluation of information gain and gain ratio, the following can be drawn:

$$\mathbf{Entropy} = \sum_{\mathbf{i=0}}^{\mathbf{n}} -\mathbf{p_i log(p_i)}$$

By dividing the trained instances according to this metric measure, information gain was employed by to assess the split metric's efficacy in decision tree construction. The gain measure of an attribute A in relation to a set of examples I is defined using the formula below:

$$\mathbf{Gain(I, A)} = \mathbf{Entropy(I)} - \sum_{\mathbf{i \in values(A)}} \left|\frac{\mathbf{I_v}}{\mathbf{I}}\right| \mathbf{Entropy(I_v)}$$

Where Iv is the subset of I for which the metric a has instance value v and value (A) denotes the set of all distinct values for metric A. Similar to this, split-info provides the following as the metric selection measure of a discrete characteristic A, relative to the specified instances I:

$$\mathbf{SplitInfo(I, A)} = -\sum_{\mathbf{i=1}}^{\mathbf{m}} \left|\frac{\mathbf{I_i}}{\mathbf{I}}\right| \mathbf{log} \left|\frac{\mathbf{I_i}}{\mathbf{I}}\right|$$

A metric A's gain ratio is specified as follows:

$$\mathbf{GainRatio} = \frac{\mathbf{Gain}(\mathbf{I}, \mathbf{A})}{\mathbf{SplitInfo}(\mathbf{I}, \mathbf{A})}$$

2. Literature Review

The improvement of body sensors has given IoT-based brilliant medical care frameworks another bearing. IT devices have been involved by these frameworks for various errands, for example, getting to patient records, recognizing extreme diseases, and distinguishing outliers in wellbeing information. Specialists and specialists use medical services information to analyze fundamental issues and to treat patients from a distance(Char, Abràmoff, & Feudtner, 2020). Machine learning approaches have been utilized to answer various exploration issues relating to medical services frameworks. To sort persistent diseases, Jain et al. (2020) utilized a two-stage crossover machine learning order framework. They asserted that the most extreme order exactness of their calculation is 98.5%. Singh et al. (2020) led a presentation examination of different machine learning calculations on different rules to distinguish cervical malignant growth. On different preparation sizes, they analyzed the accuracy and processing proficiency of various learning approaches.

Different issues with medical care frameworks are progressively being tended to utilizing IoT and other IT-empowered strategies. The troubles that Jayaraman et al. (2019) show the way that IoT could address in the medical care store network process incorporates item reviews, item deficiencies, and termination checking.They discussed how to develop and carry out these operations securely and effectively using IoT and blockchain technology. The ability to predict nurses' perceptions of their performance has been demonstrated by Ashtari et al. (2019). To validate their secure approach, they used a variety of linear regression techniques. An apriori algorithm-based model was put forth by Wang et al. (2019) to forecast lifestyle disorders and improve patient care. For their check-up, users can enter some of their fundamental information into this system. This algorithm forecasts the illness and offers recommendations accordingly.

Patients' health information is extremely important, but different people may need to access it for different purposes. Therefore, this information should be dependable and safe. Gope et al. suggested an Internet of Things-based medical care framework for proficient and secure patient checking. As indicated by Sittig et al. (2018), merchants, IT divisions, and medical care associations ought to all get a sense of ownership with the security and wellbeing of brilliant medical services frameworks. They added that the medical care framework would be more secure and further developed assuming all partners shared liabilities.

We also come across the phrase "outlier" while discussing the security of healthcare data. Outlier or anomaly data are those that drastically differ from the other data in any category. Such abnormalities might be welcomed on by malignant movement or weird sensor conduct. Wellbeing records for patients are contained in medical services information, and any progressions to this information could have disastrous outcomes(Datta, Barua, & Das, 2020). Hence, medical services related information should be liberated from irregularities. Machine learning methods can serve to precisely distinguish them. Various journalists have made machine learning-based approaches for outlier detection in medical care frameworks.

An information driven strategy for outlier detection in the patient observing framework. Assuming that there is any odd takeoff from the patients' verifiable information, a caution is created. They worked with the patient's cardiovascular information. The certifiable alarm rate given by this procedure is in the scope of 25% to 66%. Another methodology was createdto distinguish outliers in

information from medical sensors. This method is additionally incredible at spotting misleading problems(Essien A, 2020). The creators guaranteed that their methodology has a low level of bogus positive outcomes and a high pace of detection. A further machine learning-based way to deal with recognizing outliers in the medical care framework was created, with exactness ready rates going from 44% to 71%.However, by developing certain more cutting-edge techniques, the effectiveness and accuracy of these strategies can be further improved.

A few bunching and characterization based solo and managed machine learning methods can be utilized to track down outliers precisely in medical care information. The significant subjects of this work are "mean-shift," a grouping based strategy, and "backing vector machine," an order based calculation. The mean-shift is a strategy utilized in solo machine learning. It is a grouping calculation in view of thickness. The quantity of bunches isn't expressed at the start. Bunches can be circular or round, among different shapes. This calculation has a high computational intricacy yet is versatile to outliers. A directed machine learning technique is the help vector machine. It is an order based machine learning strategy with a high detection rate and minimal computational and correspondence above(J. Ha, 2021).

2.1. Objectives of the study

- To investigate how machine learning is used in the medical sector for outlier detection.
- To comprehend and assess different machine learning methods used in medicine.

3. Machine Learning

Because computers are not naturally intelligent, giving them the ability to learn like humans is a pipe dream. When it comes to doing their jobs, humans and robots differ in a few ways, one of which is intelligence. This suggests that whereas machines lack the ability to learn from their past mistakes, humans can. In actuality, they need to be programmed to adhere to certain directives. These days, computers can learn from their experiences thanks to machine learning. Traditional computational methods used to explicitly incorporate a set of programmed instructions, or were "hard coded." As opposed to the past, when PCs depended on these guidelines to settle issues, machine learning today empowers PCs to learn dynamic standards, eliminating the requirement for developers to build these principles physically. It is "Delicate customized," as the expression goes. Machine learning is a subset of man-made reasoning (AI). ML-controlled machines are more wise and independent than conventional machines. In fact, "savvy machine" is an image.It discusses the goals of machine learning.Can a machine think? is a question initially posed by Allan Turing in 1995. He developed a test known as the "Turing Test." A machine's intelligence is assessed using this test. Various definitions of machine learning exist today. For instance, machine learning is described by Arthur Samuel as "a study area that enables computers to learn without explicit programming"(J. Joseph, 2021). A further definition of machine learning provided by EthemAlpaydin is "a field for programming computers based on data samples or experience to improve a performance criterion". Machine learning refers to the search process in the space of potential representations to produce the best representation depending on the data at hand. Moreover, a hunt calculation is alluded to as a "machine" in this unique circumstance. This calculation is a mix of science \sand rationale. The overall objective of machine learning is to respond to the inquiry, "How might a PC program be created using verifiable information to take care of an issue and naturally further develop the program's presentation utilizing experience?"

truly, machine learning is an innovation for building PC calculations that gain from their environmental elements and copy human knowledge(J. S. Keertan, 2021). In machine learning, a framework is created and prepared utilizing a lot of information (a great many information tests) to handle very troublesome errands. The target of this worldview is to perform undertakings, for example, expectation, navigation, or activity without unequivocal programming.The output that is sought must be produced by this model from the inputs it receives.Humans can sometimes comprehend this model with ease. But occasionally, it resembles a black box. This implies that this model is difficult for people to comprehend. In actuality, this model simulates the process, which a machine must replicate.

3.1. ML applications in healthcare

There are numerous uses for machine learning in the medical field. It can make difficult, laborious tasks in this area easier. The development of faster processors, machine learning (ML), and access to digital health data have all made it possible to improve the healthcare system today. These new technologies increase therapeutic outcomes, lower costs, and speed up appropriate drug discovery. The major companies in the healthcare industry are currently drawn to machine learning. In general, there are three types of ML applications in the medical field:

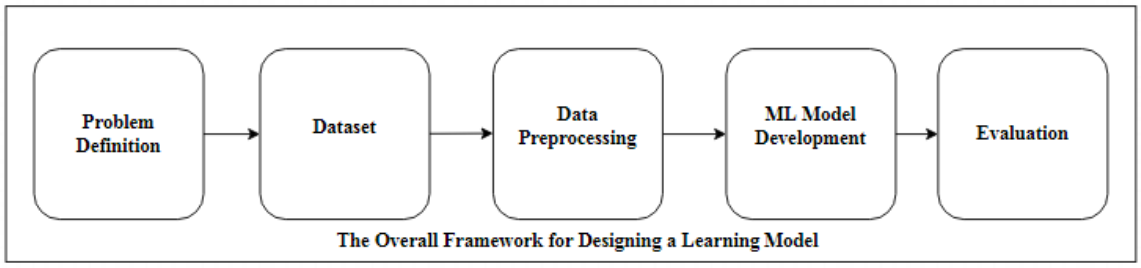
- **First Category:**Modernizing Medical Facilities These are the most direct machine learning (ML) applications utilized in the medical services area. They work on the effectiveness of existing designs. These ML-based arrangements frame specific, rule-based errands for normal applications like recreation and information checking. The order of advanced medical pictures is one of these purposes of machine learning in medical care. It expands the accuracy of customary picture handling strategies. To decide if a particular disease is available or not, radiological pictures can likewise be broke down utilizing machine learning(Jamshidi M, 2020). Furthermore, ML can be applied to assess retinal pictures and recognize patients who might be defenceless against visual dangers. For instance, the artificial intelligence and machine learning-based medical start-up Aindra. It classifies medical photos using a platform with machine learning. Its goal is to more quickly and accurately diagnose tumours.
- **Second Category:** Improving Medical Facilities Machine learning applications in this category provide structures new capabilities. They progress toward individualization. One of these ML applications is precision medicine. A kind of medical therapy centers around an individual's specific necessities relying upon their qualities (for instance, the hereditary game plan of the individual). iCarbonx is one business that is pushing toward customized medical care administrations. This utilizes huge datasets, biotechnology, and computerized reasoning.
- **Third Category:**A free medical construction is the third class of ML applications, which has of late seen development. They make ML-based models to do their tasks independently relying upon pre-laid out goals. One of the field of medical care's likely applications, for example, is the development of clinics without specialists. In this manner, we should plan for a mechanical future that relies upon AI and machine learning. To plan for the utilization of robots in clinics later on, we should now. Every single medical therapy, from diagnostics to medical procedure, will before long be performed by robots. In present day rich nations like China, Korea, and the United States, robots help specialists in the working room(Johri, Goyal, Jain, Baranwal, Kumar, & Upadhyay, 2021). Albeit this new innovation has various downsides and blemishes, it is still in its early stages and must be refined. For instance, the Mayo Clinic is edging closer to being a hospital without

physicians. They currently create its parts. However, adequate testing of these components in accordance with relevant standards should be performed. Robots are now used by doctors to facilitate surgery.

3.2. The General Framework for Designing a Learning Model in Medicine

Here, we give every one of the systems important to foster a learning model for the medical care area. Recollect that the motivation behind this part is to teach scientists on the most proficient method to foster a learning model for the medical business. To have an exhaustive cognizance of and skill about learning models, we encourage specialists to assess and lead extra exploration in this field. While building a learning model for the medical care area, we should think about five critical stages: issue definition, dataset, information pre-handling, ML model creation, and assessment. These stages are displayed in Figure 1. The segments that follow give a definite clarification of every one of these stages.

Figure: 1.Different phases for designing a learning model.



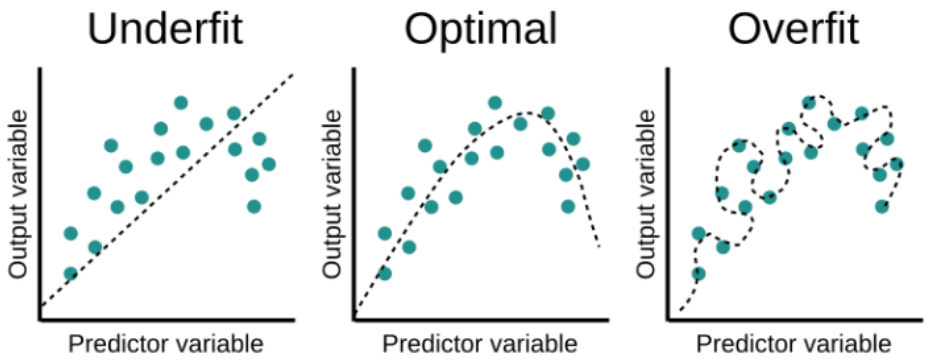
❖ **Problem Definition:** We must first provide a solution to the following query when creating a learning model for the healthcare industry: "What is the goal of constructing this learning model?" The first step in creating a usable model is to pinpoint issues and difficulties in the healthcare industry. Researchers ought to examine precisely how machine learning may be used to enhance medical care. They ought to have a look at the solutions that have already been offered in this area. Examining the data availability is an important step in the first phase(Lian W, 2020). This proposes that since there ought to be adequate information to create and test a learning model, scientists ought to know about the information sources that are as of now being used. Absence of computerized information, patient security issues, business concerns, or surprising diseases are only a couple of the variables that can add to the absence of information in the medical care area.

❖ **Dataset:** Datasets are used for training, validating, and testing learning models in the healthcare industry. Demographic data, photographs, test results, genetic data, and sensor data can all be found in healthcare databases. These data are produced or collected via a variety of platforms, including personal computers, cell phones, mobile applications, network servers, e-health records, genetic data, and wearable technology. Global linkages could be strengthened today thanks to the Internet and cloud computing. As a result, data accessibility has increased. It is important to lay out the right methodology for assessing the learning model prior to creating it for the medical care business since it isn't adequate for machine learning for the creator to guarantee that its learning model has a superior presentation and is entirely attractive. The focal point of ML-based models is information.As a result, individuals could experience an issue known as over fitting or under fitting. Over fitting and under fitting should be traded off in an effective learning model. This implies that it must have a suitable bias and a suitable variance(Liu H, 2021). When we create a relatively basic

learning model in comparison to the difficulty of the issue and the size of the dataset, under fitting happens. Both on training sets and testing sets, this learning model performs poorly. This indicates that it is very biased. In contrast, over fitting also happens when the learning model is extremely complicated and has many parameters in comparison to the difficulty of the issue and the amount of the dataset. Specifically, this model performs well on the training dataset while performing poorly on the testing set. In this case, there is a significant variation. Low bias and low variance are typical characteristics of an effective learning model.

Situation	Training Error Rate	Testing Error Rate
Under fitting	High	High
Over fitting	Low	High
Balanced	low	Low

Figure: 2. Over fitting or under fitting description.



❖ **Data Pre-Processing:** Data preparation is one of the most difficult problems to solve when creating a learning model for the healthcare industry since a machine learning model needs high-quality data to produce better training results and more accurate results. The process of evaluating noisy data, missing values, duplicate data, and contradicting data is known as data pre-processing. Before building the learning model, this technique is meant to improve the database's quality. In order to estimate missing values or filter outliers, data pre-processing may be necessary(M. Alam, 2020). Some data reduction techniques, like feature selection or feature extraction, can be applied if data also have high dimensions. The best subset of features is chosen through feature selection. On the other hand, feature extraction uses the initial dataset to find a new dataset with fewer dimensions.

❖ **ML Model Development:** Building a learning model for the medical services area requires thought of the information base size, kind of learning plan, and model surmising time. We base a learning model's intricacy on the size of the data set to keep away from overfitting or underfitting. It's vital to consider a learning model's preparation span. More defined learning models, notwithstanding, can bring about additional exact results. These models need additional opportunity to prepare and lead more computational tasks in this present circumstance. Thus, they can't be utilized for continuous applications. Building an inclining model is consequently more qualified for lightweight development(Nordlinger, Villani, & Rus, 2020). It is fundamental to consider the kind

of learning plan while creating machine learning models. The four primary sorts of learning are unaided learning, semi-regulated learning, administered learning, and support learning.

❖ **Evaluation:** Evaluation of a machine learning-based system requires running multiple procedures to look for differences between the system's current behaviour and the expected behaviour. The proper evaluations should be conducted after developing a learning model for the healthcare sector to determine whether the model is appropriate for deployment in real-world settings. A number of scales are used by designers to assess the learning model's efficacy.Its strengths and faults are determined by this evaluation. Additionally, we must reevaluate the learning model's effectiveness after deploying it in actual contexts in order to assess how it behaves when interacting with actual users. A machine learning framework can be assessed in different ways, for example, by surveying the information used to make the last learning model, the learning calculations used to make the last model, and the last model's exhibition.

4. Outlier Detection Using Machine Learning Algorithm

4.1. Feature engineering

One of the feature engineering features was the discrepancy between the patient's data's current initial value and its ultimate value. The second feature measured the discrepancy between the most recent value of the medical data and the average of the previous five values, and the last feature was a machine learning clustering feature built on the aforementioned data and based on 100 groups. If the data is not normalised, the models perform less well.

4.2. Simulated data performance

Each algorithm was tested and tuned using a 0.5% outlier dataset. Create an 80:20 split of the data between the training and test sets. Sample parameters were given in the model using a sequence of connected loops that changed each pertinent parameter. Through this type of adjustment, we were able to assess the importance of characteristics to the model and, as a result, remove any that were unnecessary(P. V. Amoli, 2021).

The suggested clustering approach is a straightforward and well-liked one. To identify a point as an outlier, it considers distance and the minimum required number of points in each cluster. Thus, the robust algorithm makes two predictions: first, it determines whether the point is an outlier. Check all clusters besides the outlier cluster to improve the predictions. The default calculating method will be the Euclidean distance function.

The proximity of an object is indirectly detected by a Euclidean distance-based approach, which is defined by a specific radius. One might think of a distance based on the threshold as an object's neighbourhood. We can identify an object's neighbours for each object o. Let the threshold fraction be (01) and the threshold distance be r(r>0).

$$\mathbf{dist} = \frac{\|\mathbf{o'dist(o,o')} \leq \mathbf{r}\|}{\|\mathbf{D}\|} \leq \emptyset \mathbf{(1)}$$

The second strategy requires O (n2) time,

- A weight-based outlier detection technique is used to assess an object's density and the density of its neighbours.
- Numerous real-world data sets have intricate structures. Local communities are better at measuring item outliers than global data.

The methods for density-based outlier detection focus on the densities of nearby sites. The distance between an object and KNN is called dist k(o). The k-distance neighbourhood of o is closer to dist k when viewed as a whole object from o's distance (o).

$$\mathbf{N_k(o) = [o|o' \in D, (o, o') \leq dist_k(o)]} \quad \textbf{(2)}$$

Determine the typical distance between each object in Nk(o) and o. The distance calculation may encounter extremely substantial numerical differences if o has very close neighbours o' such that dist (o, o') is a very short distance.Therefore, normalisation techniques were used to resolve the current problems.

$$\mathbf{reachdistk(o, o') = max[distk(o), dist(o, o')]} \quad \textbf{(3)}$$

k is a user-specified parameter that specifies the smallest neighbourhood in which to check an object's local density. An object's local density is,

$$\mathbf{lrd(o) = \frac{\|N_k(o)\|}{\Sigma_{o' \in N_k(o) reachdist_k(o,o')}}} \quad \textbf{(4)}$$

To determine the extent to which an object is considered an outlier, we compute its density for local reachability and compare it to its neighbours. The aforementioned algorithm functions in the same way as other algorithms. There are, however, variances in effectiveness and suitable datasets with different data sizes.

$$\mathbf{LOF_K(o) = \frac{\Sigma o' \in N_k(o) \frac{lrd_k(o')}{lrd_k(o)}}{\|N_k(o)\|}} \textbf{(5)}$$

3.3. Proposed algorithm

Begin
Step1: from cluster1 import ML-A
Step2: outlier1_detection = ML-A (eps = .2
Step3: measurement=”euclidean”,
Step4: min_samples = 5, n_jobs = -1)
Step5: clusters = outlier_detection.fit_predict (num2) end

5. Result and Analysis

Medical data analysis is crucial in real-world scenarios to produce reliable results. Predict the real-time results of this study to find outliers. Before creating a prediction, look into the past state of the available data. After a primary task, this algorithm gives decent results with little sensitivity. Real-time medical data are employed with an outlier detection model. In that it involves pre-handling the information, preparing the information, testing the information utilizing the preparation information, approving the information, and conveying exact outcomes, the preparation and testing technique is like the information mining procedure(Pattnayak & Jena, 2021).

With datasets that have a 2.5% outlier rate, the given models perform well. The upper finish of the medical information range is where the LOF searches for outliers. From the introduced information at radio recurrence (RF) and moderate recurrence, this study expected to distinguish outliers (IF). The model prepared on 2.5% of the dataset performed well, similar as partner was prepared on the 0.5% outlier dataset. It is significant to feature that the ideal balance between the hit and phony problem rates not set in stone by the undertaking related punishments of the outlier acknowledgment.The classification of patterns has not changed as a result of different training dataset results. This position's goal is to evaluate the algorithm's prospective performance based on outliers. visualised the model's presentation using fresh, real-world medical data samples, took

notice of differences between different models, and came to the conclusion that the model should display itself differently depending on the data it was trained on.

We use this kind of approaches in our research, which is grounded in real-world considerations, when algorithm consumption is too high and it supports parallel processing. The processing and distribution of data in computing resources are supported by the algorithm's throughput. The current study was able to validate medical data through the implementation of a classification algorithm for outlier detection(R. Vijaya Kumar Reddy and U. Ravi Babu, 2020). We compare various techniques for finding outliers in Table 1 below. The suggested innovative strategy outperforms conventional ways among these techniques.

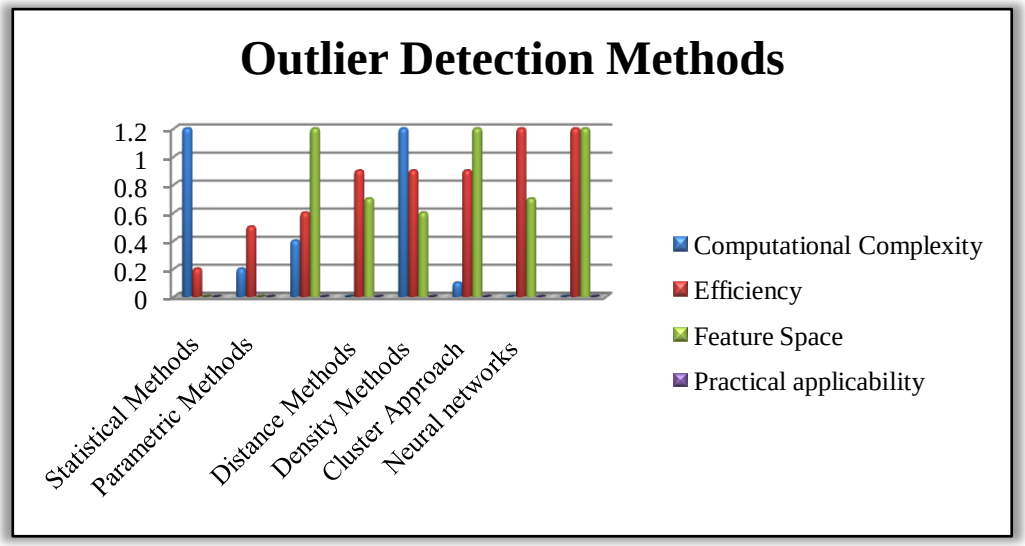
Figure 5 is a graph that compares various techniques for finding outliers. The suggested innovative strategy outperforms conventional ways among these techniques. In the presence of feature space and efficiency, the approach I've given produces better outcomes when compared to more methods.

The machine learning (ML) approach that has been proposed has very little computational complexity. Due to an AI reliance factor subject to the size of the information and the model of the information, the viable materialness isn't referenced in this occurrence. Here is an examination of a few specialized results. Our suggested outcomes state each of them separately. For our research, we used a dataset that was acquired from the Kaggle website.

Table: 1. Comparing different outlier detection techniques

Methods	**Computational Complexity**	**Efficiency**	**Feature Space**	**Practical applicability**
Statistical Methods	High Complex	Low	Single Variable	Statistical data
Parametric Methods	Low Complex	Moderate	Single Variable	Prior knowledge of data sets
Non Parametric methods	Low Complex	Moderate	Single / Multi Variable	profile of the data required
Distance Methods	Nil	High	Multi Variable	Relation of individual points
Density Methods	Complex	High	Multi Variable	Relation of points and nearest neighbour too
Cluster Approach	Low Complex	High	Single / Multi Variable	clustering of similar data
Neural networks	Nil	Very High	Multi Variable	Simple training data
Proposed ML approach	Nil	Very High	Single / Multi Variable	Based on data size

Figure: 3.Graph representation of outlier detection methods



6. Conclusion

Healthcare data is each item's actual data when establishing outlier indicators. After noise processing, the data are extremely readily available and the outliers themselves are very informative. The hospital's healthcare quality can be understood from the standpoint of data outliers using outlier identification based on these data. The utilization of machine learning in the medical services industry creates the best results as well as diminishes the responsibility.This algorithm might solve the problems and uncover fresh information for the advancement of medicine in the healthcare sector. This research suggests a novel approach for identifying outliers in various medical datasets(Rincy N. T, 2021). Taking into account that medical data analyses health issues the suggested method relies on both supervised and unsupervised learning to function. The outliers in medical data are found using this approach. The efficiency of using local and global data factors to quickly identify outliers in medical data. Whatever the case, the model in this scenario was created by them and tested using medical data. The statistical findings demonstrate that the outlier recognition technique based on machine learning offered the highest accuracy.

We saw ML-based strategies in medical care in this exploration. To do this, we previously gave a brief clarification of AI and examined its utilization in medical care.Then, we presented a wide construction for making ML-based clinical models.The system we utilized in this paper enjoys a few observable benefits as well as disservices. We have chosen to utilize an openly available dataset with a predetermined number of sources of info and cases. The information is coordinated such that simplifies it for individuals with preparing in clinical fields to draw correlations between notable measurements and state of the art ML procedures.

References

1. A. Nisioti, A. Mylonas, P. D. Yoo, and V. Katos, “From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods,” IEEE Communications Surveys Tutorials, vol. 20, no. 4, pp. 3369-3388, 2021, doi: 10.1109/COMST.2018.2854724.
2. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict covid-19 infection. Chaos SolitFract. 2020;140

3. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. Algorithms. 2020;13(10):249.
4. Char, D.S.; Abràmoff, M.D.; Feudtner, C. Identifying ethical considerations for machine learning healthcare applications. Am. J. Bioeth. 2020, 20, 7–17.
5. Datta, S.; Barua, R.; Das, J. Application of artificial intelligence in modern healthcare system. In Alginatesrecent Uses of This Natural Polymer; IntechOpen: Rijeka, Croatia, 2020.
6. Essien A, Petrounias I, Sampaio P, Sampaio S. A deep-learning model for urban traffic flow prediction with traffic events mined from twitter. In: World Wide Web, 2020: 1–24 .
7. J. Ha, S. Seok and J.-S. Lee, "A precise ranking method for outlier detection," Information Sciences, vol. 324, pp. 88-107, Dec. 2021, doi: 10.1016/j.ins.2015.06.030.
8. J. Joseph, "How to detect outliers using parametric and non-parametric methods: Part I", 2021.
9. J. S. Keertan, Y. Nagasai and S. Shaik, "Machine Learning Algorithms for Oil Price Prediction," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.. 8, pp. 958-963, June. 2021.
10. Jain D. Singh V. (2020). A Novel Hybrid Approach for Chronic Disease Classification. International Journal of Healthcare Information Systems and Informatics (IJHISI), 15(1).
11. Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, La Spada L, Mirmozafari M, Dehghani M, et al. Artificial intelligence and covid-19: deep learning approaches for diagnosis and treatment. IEEE Access. 2020;8:109581–95.
12. Johri, S.; Goyal, M.; Jain, S.; Baranwal, M.; Kumar, V.; Upadhyay, R. A novel machine learning-based analytical framework for automatic detection of COVID-19 using chest X-ray images. Int. J. Imaging Syst. Technol. 2021, 31, 1105–1119.
13. Lian W, Nie G, Jia B, Shi D, Fan Q and Liang Y, "An Intrusion Detection Method based on Decision Tree-Recursive Feature Elimination in Ensemble Learning", Proc. of Mathematical Problems in Engineering,2020.
14. Liu H, Lang B, "Machine Learning and Deep Learning methods for Intrusion Detection System: A Survey", Applied Sciences, MDPI, 2021.
15. M. Alam, "Statistical techniques for anomaly detection", September 2020.
16. Nordlinger, B.; Villani, C.; Rus, D. Healthcare and Artificial Intelligence; Springer Nature: Cham, Switzerland, 2020.
17. P. V. Amoli, T. Hamalainen, G. David, M. Zolotukhin, and M. Mirzamohammad, "Unsupervised network intrusion detection systems for zero-day fast-spreading attacks and botnets," International Journal of Digital Content Technology and Its Applications, vol. 10, no. 2, pp. 1-13, 2021.
18. Pattnayak, P.; Jena, O.P. Innovation on Machine Learning in Healthcare Services—An Introduction. Mach. Learn. Healthc. Appl. 2021, 1–15
19. R. Vijaya Kumar Reddy and U. Ravi Babu, "A Review on Classification Techniques in Machine Learning," International Journal of Advance Research in Science And Engineering, vol. 7, no. 3, March 2020.
20. Reig, B.; Heacock, L.; Geras, K.J.; Moy, L. Machine learning in breast mri. J. Magn. Reson. Imaging 2020, 52, 998–1018.

21. Rincy N. T, Gupta Roopam, “Design and Development of an efficient Network Intrusion Detection System using Machine Learning Techniques”, Wireless Communications and Mobile Computing, 2021.
22. S. Shaik and U. Ravibabu, “Classification of EMG Signal Analysis based on Curvelet Transform and Random Forest tree Method,” Journal of Theoretical and Applied Information Technology (JATIT), vol. 95, no. 24, pp. 6856-6866, Dec. 2021.
23. Shenfield A, Day D, Ayesh A, “Intelligent Intrusion Detection Systems using Artificial neural networks”, The Korean Inst. Of Communications and Information Sciences, ICT Express 4, pp. 95-99, 2020.
24. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif. Intell. Med. 2020, 104, 101822.
25. Yousefpoor, E.; Barati, H.; Barati, A. A hierarchical secure data aggregation method using the dragonfly algorithm in wireless sensor networks. Peer-to-Peer Netw. Appl. 2021, 1–26.
26. S. Srivastava and R. Kumar, "Indirect method to measure software quality using CK-OO suite," 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), 2013, pp. 47-51, doi: 10.1109/ISSP.2013.6526872.
27. Ram Kumar, Gunja Varshney , Tourism Crisis Evaluation Using Fuzzy Artificial Neural network, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011
28. Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, “A Survey Paper on Altered Fingerprint Identification & Classification” International Journal of Electronics Communication and Computer Engineering Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209
29. Kumar, R., Singh, J.P., Srivastava, G. (2014). Altered Fingerprint Identification and Classification Using SP Detection and Fuzzy Classification. In: , et al. Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1602-5_139
30. Gite S.N, Dharmadhikari D.D, Ram Kumar,” Educational Decision Making Based On GIS” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012.
31. Ram Kumar, Sarvesh Kumar, Kolte V. S.,” A Model for Intrusion Detection Based on Undefined Distance”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011
32. Vibhor Mahajan, Ashutosh Dwivedi, Sairaj Kulkarni,Md Abdullah Ali, Ram Kumar Solanki,” Face Mask Detection Using Machine Learning”, International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022
33. Kumar, Ram and Sonaje, Vaibhav P and Jadhav, Vandana and Kolpyakwar, Anirudha Anil and Ranjan, Mritunjay K and Solunke, Hiralal and Ghonge, Mangesh and Ghonge, Mangesh, Internet Of Things Security For Industrial Applications Using Computational Intelligence (August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187998 or http://dx.doi.org/10.2139/ssrn.4187998
34. Kumar, Ram and Aher, Pushpalata and Zope, Sharmila and Patil, Nisha and Taskar, Avinash and Kale, Sunil M and Gadekar, Amit R, Intelligent Chat-Bot Using AI for Medical Care

(August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187948 or http://dx.doi.org/10.2139/ssrn.4187948

35. Kumar, Ram and Patil, Manoj, Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies (July 22, 2022). Available at SSRN: https://ssrn.com/abstract=4182372
36. Ram Kumar, Manoj Eknath Patil ," Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies", Turkish Journal of Computer and Mathematics Education ,Vol.13 No.3(2022), 6168-6174.
37. Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, "A Survey Paper on Altered Fingerprint Identification & Classification" International Journal of Electronics Communication and Computer Engineering ,Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209.