Outlier Identification Based on Machine Learning for Medical Equipment

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

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

[2] Research Scholar, 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: 11193-11206
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 15 September 2022
Revised: 25 October 2022
Accepted: 14 November 2022
Publication: 21 December 2022

Abstract

In real-time medical databases, feature selection methods are crucial. The majority of medical databases have high dimensionality and enormous amounts of data, making it challenging to identify a crucial key feature using conventional feature sub-set selection methods. Additionally, due to the high data size and feature space, typical medical data filtering algorithms fall short of identifying the crucial outliers. The use of machine learning in the healthcare industry not only produces the best outcomes but also lessens the workload. This algorithm might solve the problems and uncover fresh information for the advancement of medicine in the healthcare sector. The new approach for locating outliers using various datasets is proposed in this research. Taking into account that medical data analyse both a health issue and an activity. Based on both supervised and unsupervised learning, the suggested method operates. 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 cleaning procedure using all of the dataset's properties for similarity operations. Various medical datasets with built-in experimentation are used. The statistical results show that the outlier identification technique based on machine learning has the highest level of accuracy.

Keywords: Machine Learning, Outlier identification, Medical equipment, Outlier detection in medical.

---

1. Introduction

Outlier detection continues to be a critical and broad review field in medicine because of its continuous use in various applications. Analysts can gain essential information that assists them with making better information decisions by locating outliers.Additionally, identifying outliers yields valuable information that can be put to use in a number of health diagnoses. An outlier is often conceived of as a data point that differs significantly from other data points or that does not follow the expected regular pattern of the phenomenon it represents, though a specific definition is difficult to provide.

In contrast to other data mining techniques, outlier recognition focuses on discovering patterns that occur infrequently. A dataset's large inconsistencies are the source of an outlier. Since outliers can provide bare patterns and priceless information about a dataset, outlier detection is important in the eyes of the truth (A. Lal). The areas of outlier discovery, credit card fraud detection, network intrusion detection, criminal detection, medical analysis, and uncommon part detection in image processing are all covered by current research.

Unsupervised outlier detection often includes approaches based on distance, density, and distribution. This method reveals that every data point is the result of a certain mathematical model, but outliers reject this kind of model. Local dataset information differs from global parameters. Breunig et al. originally discussed the density-based approach. Every data point is given a local outlier factor based on their local point density. An outlier is a data point having a local (LOF) value that is significantly above the mean. The approaches based on clustering are unsupervised, need no labelled training data, and only sometimes contribute to outlier discovery.

A limited number of objects are tagged as outliers to a certain class in many real-world applications, but the majority of data are unlabeled. To significantly increase the effectiveness of outlier detection, some basic information is needed. In order to handle this type of issue, which has been taken into account of a well-liked track of outlier finding, semi supervised approaches for outlier recognition have been created.

1.1. Outliers Detection in Medicine

In medicine, outlier detection is crucial because these data can provide essential information that is frequently crucial. Outlier detection algorithms are used in healthcare databases to identify unusual patterns in patient records that may contain useful information, such as symptoms of a new disease (Banerjee S, 2021).

The vocabulary used in diverse studies uses a variety of definitions for outliers. According to Barnett and Lewis' definition of an outlier, an observation, or subsets of observations, that stand out from the rest of the data. An outlier is a part that strays from the ordinary gathering of information that it is a piece of. All things considered, an outlier is generally a piece of a gathering. Since a component supposedly is an outlier when contrasted with a norm, it very well may be contrasted with the standard X as opposed to the standard Y.

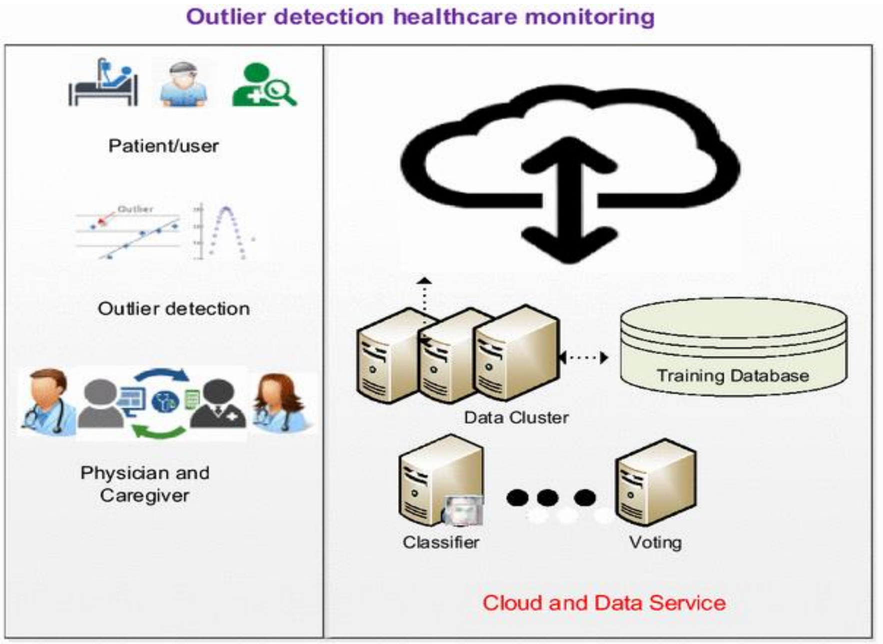
1.1.1. Techniques

The outliers detection strategies under discussion in this work are:

- **Statistical:** After measurable cycles have fitted a factual model, as often as possible for typical way of behaving, to the given information, a factual inference test is utilized to determine whether an unseen event compares to the model. Outliers are occasions that have a low likelihood of being delivered by the learned model, as determined by the test measurement.
- **Clustering:** Clustering is used to cluster together comparable pieces of data. Several clustering-based outlier identification algorithms have been developed, despite the fact that the two concepts initially appear to be fundamentally unrelated.
- **Classification:** Using this procedure, a test instance is characterized into one of the classes by a learning model that was trained on a progression of marked information instances. Classification-based anomaly detection methods categorise test cases as normal or anomalous using a classifier developed during the training phase using the available labelled training data, much like other two-phase operations.

- **Nearest Neighbour:** A distance or similarity measure between two data instances that must be defined and can be determined in many methods. This approach's underlying algorithms can be generally divided into two categories: those that compute each data instance's anomaly score based on its relative density and those that compute the distance between each data instance and its nearest neighbour.
- **Mixture Models:** Mixture models can explain a variety of data features by combining a limited or infinite number of components, some of which may have distinct distributional types. A mixture model in statistics uses a mixture distribution as the basis for a probabilistic density estimation model. A mixed model might be compared to unsupervised learning or clustering.
- **Spectral:** By combining attributes that account for the majority of the data's variability, try to get an approximation of the data. This method specifies the subspaces where the anomalous instances can be quickly located.

Figure: 1. Overall scenario for outlier detection in Healthcare



2. Literature Review

In healthcare informatics and medical diagnostics, the usage of outlier identification tools is vital. Atypical symptoms noted in healthcare records or test results could indicate a health problem. The identification of unexpected data should ideally be able to differentiate between instrumentation or recording errors and clinically significant changes in the patient's condition to enable quick medical action in the latter scenario. The data typically comprises of records with a variety of features, such as patient age, weight, physiological signals (such as the EKG), vital signs (such as heart rate), blood test results, and medical imaging data, according to Clifton et al. (2011) and Lin et al. (2005).

Z. Cömert, B. Ergen, and M. Toaçar, 2020) discovered a feature selection technique, known as consistent-based feature selection. It served as an important indicator for several device selection strategies. Thus, the proposed strategy produced a higher reduction in functionality and a better accuracy result than the wrapper approach. (Y.-C. Chen and C.-F. Tsai, 2021) offered a feature selection method that makes use of the mining law of association and knowledge development. The properties in question were located using the Apriori method.To remove redundant and dated characteristics from the dataset, Knowledge Gain was utilised.

2021 (P. Wiesaw) examined the discriminatory selection feature algorithm using knowledge theory as it failed to distinguish the continuous and discriminating dataset characteristics. Additionally, they put up an approach for choosing features from the analysis. This algorithm includes a definition of an entropy breakpoint. Utilizing a variety of real-world datasets, this approach has been tested. According to the experiment's findings, the method exhibited a high degree of computational complexity and a low degree of prediction precision.

2.1. Outlier Detection Methods

Supervised, unsupervised, and semi-supervised outlier identification algorithms fall under these three categories. However, there are other classifications for unsupervised outlier identification methods. Numerous unsupervised outlier identification methods have already been introduced. The above strategies can be partitioned into measurable dissemination based, clustering-based, thickness based, and model-based approaches (Beam A, 2021).

2.1.1. Supervised Outlier Detection Methods

The class objects that have been given the normal or abnormal categorization make up the dataset used in this function. The availability of a training dataset with labelled cases for both the normal and outlier classes is a prerequisite for techniques presented in supervised mode. Creating a predictive model for normal vs. outlier classes is the standard procedure in such cases. The class to which any unclassified item of data belongs is ascertained by comparing it to the model. The two main problems with supervised outlier detection are as follows. First, the training data contains fewer anomalous examples than typical examples. Second, it is frequently difficult to create labels that are accurate and representative, particularly for the outlier class. To create a set of labelled training data, several strategies have been put forth by Theiler et al. (2003), Abe et al. (2006), and Steinwart et al (2005). These techniques include adding fictitious outliers to an ongoing data collection.

2.1.2. Unsupervised Outlier Detection Methods

It is the method where the dataset's class distribution is unknown beforehand. These days, a lot of people use this tactic. Techniques in unsupervised mode are the most popular since they don't require training data. The implicit assumption made by the approaches in this category is that the test data comprises a far higher percentage of typical cases than outliers. If this assumption is erroneous, then these processes frequently cause false alarms. Unsupervised outlier identification techniques can be categorised into three types of methodologies: statistical, proximity-based, and model-based. The three categories of proximity techniques, however, are cluster-based, distance-based, and density-based (C. Huang et al., 2020).

2.1.3. Semi-Supervised Outlier Detection Methods

It is expected for semi-supervised algorithms that the training set contains only labelled cases for the typical class. They are more frequently used than supervised approaches since they do not require labels for the outlier class. For instance, in the difficult to model field of spaceship fault detection, an outlier state would indicate an accident. Making a model for the class that depicts typical behaviour and using that model to find outliers in the test data is the common method employed in such strategies. Solely a small number of outlier detection methods rely on access to only outlier data for training (Forrest et al. 1999, Dasgupta et al. 2000 and Dasgupta et al. 2002). Since it is hard to find a training informational index that contains each conceivable unusual way of behaving that could happen in the information, such methods are not habitually utilized.

2.2. Objectives of the study

- To increase comprehension of the techniques used to identify outliers in medical data.
- To investigate the application of ML to medical databases to find outliers.

3. Evaluation Techniques, Tools, and Datasets For Outlier Detection Problems

3.1. Evaluation Methods

Over the years, a variety of outlier identification algorithms have been put forth; nonetheless, evaluating these techniques remains a significant difficulty in data mining research. To address this issue, various approaches have been suggested. Without conducting a full analysis from a wider perspective, some academics have asserted that their method beats other ways as the number of outlier identification algorithms has grown over time (D. Hendrycks, 2021). Since there are as of now no settled procedures for evaluating various calculations, scientists have regarded this as an open examination bearing.

Recently, other studies have concentrated on evaluating OD techniques, such as distance-based, ensemble-based, and unsupervised techniques.

There is a lot of research on distance-based methods from the past ten years, and assessing these methods is highly important. However, comparing these methods to one another is not without its difficulties. The effectiveness and efficiency of outlier detection problems are what most researchers and practitioners are most interested in. For instance, comparing the efficacy of various methods can be challenging because performance will vary depending on the amount of data used, the dimensionality of the dataset, the parameter selection, and other implementation-specific aspects.

Orair et al. zeroed in on evaluating a few distance-based procedures for outlier detection approaches in request to obtain some supportive direction for developing the best outlier detection calculations. A few as of late proposed improvement strategies were exposed to a factorial trial, and an outlier detection strategy was executed. Assessing the benefits and disservices of specific advancement procedures is finished (Domingues, Filippone, Michiardi, & Zouaoui, 2020). The results of the factorial analysis can be utilized to gain huge and fascinating insights. For instance, they found that particular streamlining combinations function admirably for both genuine world and engineered informational collections. In any case, none of the streamlining combinations can profess to be better than the others for a wide range of information. Three methodologies of streamlining were suggested by the creators.

i. *Approximate Nearest Neighbour Search (ANNS)*
ii. *Pruning:* is a pre-processing step that many algorithms employ to help with data point partitioning or clustering. The authors' suggested trimming strategies are:
   - Pruning partitions during the search for Neighbours (PPSN).
   - Pruning partitions during the search for outliers (PPSO).

iii. *Ranking:* The primary goal is to increase the ANNS pruning rule's effectiveness. There are two types of optimization strategies, and they are as follows:

- Ranking Objects Candidates for Neighbours (ROCN)
- Ranking Objects Candidates for the Outlier (ROCO)

3.2. Tools For Outlier Detection

Many technologies and datasets have been utilised in outlier detection. Here, we provide some widely used databases for outlier discovery as well as various outlier detection techniques. The creation of numerous software tools, including those listed below, has been prompted by the frequency of outlier detection in industrial applications.

- ❖ **Scikit-learn Outlier Detection:** Some machine learning methods for outlier detection issues are available through the scikit-learn project. Some of the algorithms it uses include LOF and Isolation Forest.
- ❖ **Python Outlier Detection (PyOD):**In multivariate data, outliers are found using PyOD. A scalable Python programme called new deep learning and outlier ensembles models have both used it in research and commercial enterprises.
- ❖ **Environment for Developing KDD-Applications Supported by Index-Structures (ELKI):** An assortment of data mining techniques, including OD algorithms, are offered via the open source data mining algorithm ELKI. It enables the quick, impartial evaluation and benchmarking of OD methods. Java was used to write it.
- ❖ **Rapid Miner:** Numerous well-known unsupervised outlier identification techniques, including LOF, COF, LOCI, and LoOP, are included in the tool's extension.
- ❖ **MATLAB:** Additional outlier identification strategies are supported by MATLAB and perform. You can use to implement algorithms because it is user-friendly, MATLAB.
- ❖ **Massive Online Analysis (MOA) tool:** The open source MOA structure gives a determination of information stream mining strategies. It involves explicit devices for assessment as well as a couple of strategies for locating outliers in view of distance, like COD, ACOD, Abstract C, and MCOD.

3.3. Datasets For Outlier Detection

Outlier detection strategies have been utilized to different information types, including normal and high-layered informational collections, streaming datasets, network information, uncertain information, and time series information. In the examination on outlier identification, two classes of information are generally thought of and expected for assessing the viability of the calculations (Esteva A, 2020). They are both engineered and genuine world datasets. The freely available information bases contain this present reality datasets. The following are the absolute most notable and pragmatic information bases with genuine world datasets for outlier detection:

- ❖ **The UCI repository:** Various OD approaches utilize the UCI archive, which has many transparently available informational indexes, to survey the adequacy of the calculations. However, most of these datasets are made for order procedures. Pre-processing the datasets is commonly used in outlier detection settings.The rest are regarded as the regular objects, while the outliers represent those belonging to the minor class.
- ❖ **Outlier Detection Datasets (ODDS):**Chances gives open admittance to various datasets utilized exclusively for outlier detection, rather than the UCI store. Just a small bunch of the a

wide range of kinds of datasets that have been assembled include multi-faceted datasets, time series univariate and multivariate datasets, and time series diagram datasets.

- **ELKI Outlier Datasets:** Both several data sets for evaluating OD approaches and a collection of data sets for outlier detection are available from ELKI. These data sets are used to examine how various OD techniques and parameters work.
- **Unsupervised Anomaly Detection Data verse:** These datasets are utilized to assess unaided outlier identification calculations by comparing the outcomes to the standards. Most of the informational indexes are from directed machine learning datasets, which are gathered from various sources. In light of protection and security concerns, certifiable informational collections contain a great deal of information that isn't accessible to the overall population.

4. Research Methodology

Essentially, this is a descriptive or analytical paper. The data was gathered from a variety of books, scholarly articles, newspapers, journals, and trustworthy websites. The theoretical study has been supported by a few maps, illustrations, and graphs. The secondary data sources on which this study is based. Past field studies, pertinent books, journals, census data, and reports make up the secondary sources.

In this preliminary review, the following variables are analysed:

- **Outliers detection techniques: S**tatistics, closest neighbour, classification, mixture models, and spectral.
- **Data type: T**ypes of data used with each technique, such as photographs, patient data, and bio signal data
- **Medical domain:** Which describes the field of medicine or medical expertise used in the study is used.
- **Clinical stage: P**hase clinical describes how the study is used to make diagnoses, predict results of treatments, or evaluate performance.

Moreover, insights regarding the sort of information utilized — essential, auxiliary, or recreated — the country wherein the review was directed, and the year it was distributed — were assembled.

5. Result and discussion

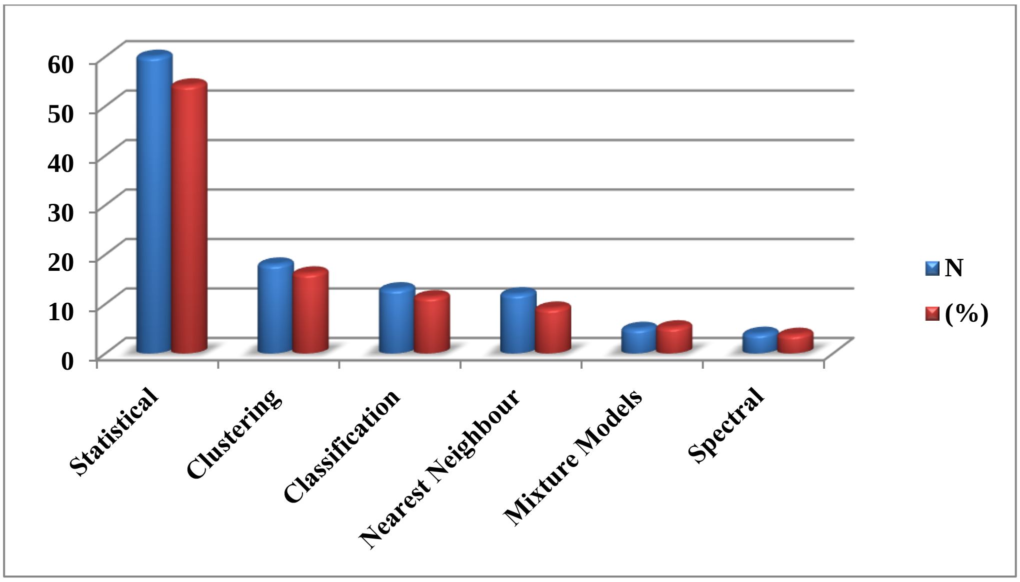
All outlier detection methods taken into consideration for this evaluation are listed in table 1. 112 techniques from the 80 publications under analysis were found and grouped into 6 categories.

Table: 1. List of outlier detection techniques

Type	N	(%)
Statistical	60	54.2
Clustering	18	16.2
Classification	13	11.4

Nearest Neighbour	12	9.10
Mixture Models	5	5.3
Spectral	4	3.8
Total	112	100.0

Figure: 2. Chart of outlier detection techniques

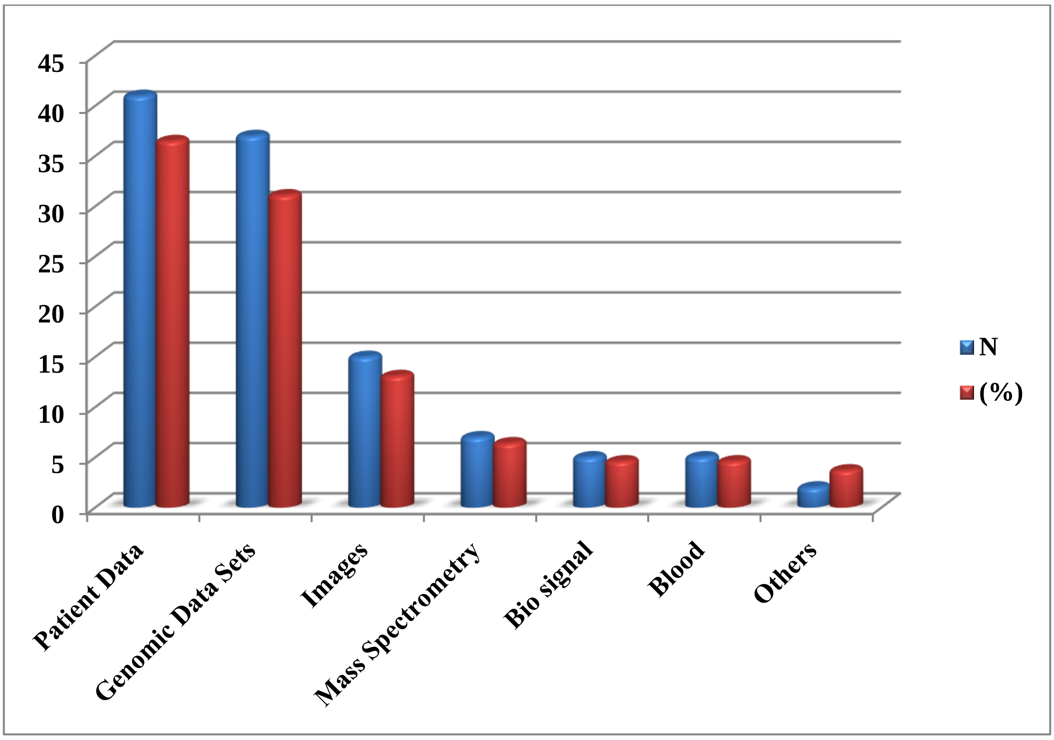


The more prevalent data types in this investigation are listed in Table 2. Patient data (36.5%), genomic data sets (31.1%), and pictures (13.1%), which comprise, for example, two- and three-dimensional photographs, electroencephalograms, magnetic resonance imaging, and ultrasonography, are listed in that order.

Table: 2. Data Type

Type	**N**	**(%)**
Patient Data	41	36.5
Genomic Data Sets	37	31.1
Images	15	13.1
Mass Spectrometry	7	6.4
Bio signal	5	4.6
Blood	5	4.6
Others	2	3.7
Total	112	100.0

Figure: 3. Chart of Data Types

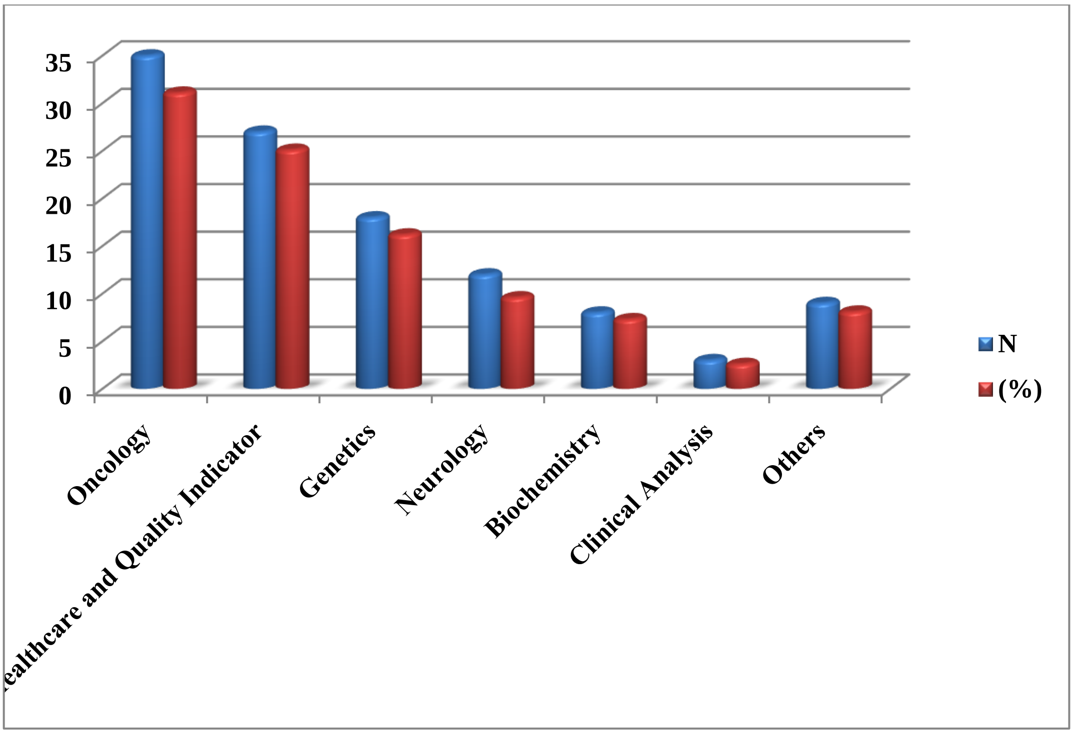


Data from the medical domain are included in Table 3. Oncology (31.1%) and genetics (16.2%) are the medical specialities that use outlier detection methods more frequently than others. When compared to medical specialties, research solely focused on administrative medical data, or "healthcare and quality indicators," are rated second (25,1%).

Table: 3. Medical Domain

Specialties	N	(%)
Oncology	35	31.1
Healthcare and Quality Indicator	27	25.1
Genetics	18	16.2
Neurology	12	9.6
Biochemistry	8	7.3
Clinical Analysis	3	2.6
Others	9	8.1
Total	112	100.0

Figure: 4. Chart of Medical Domain

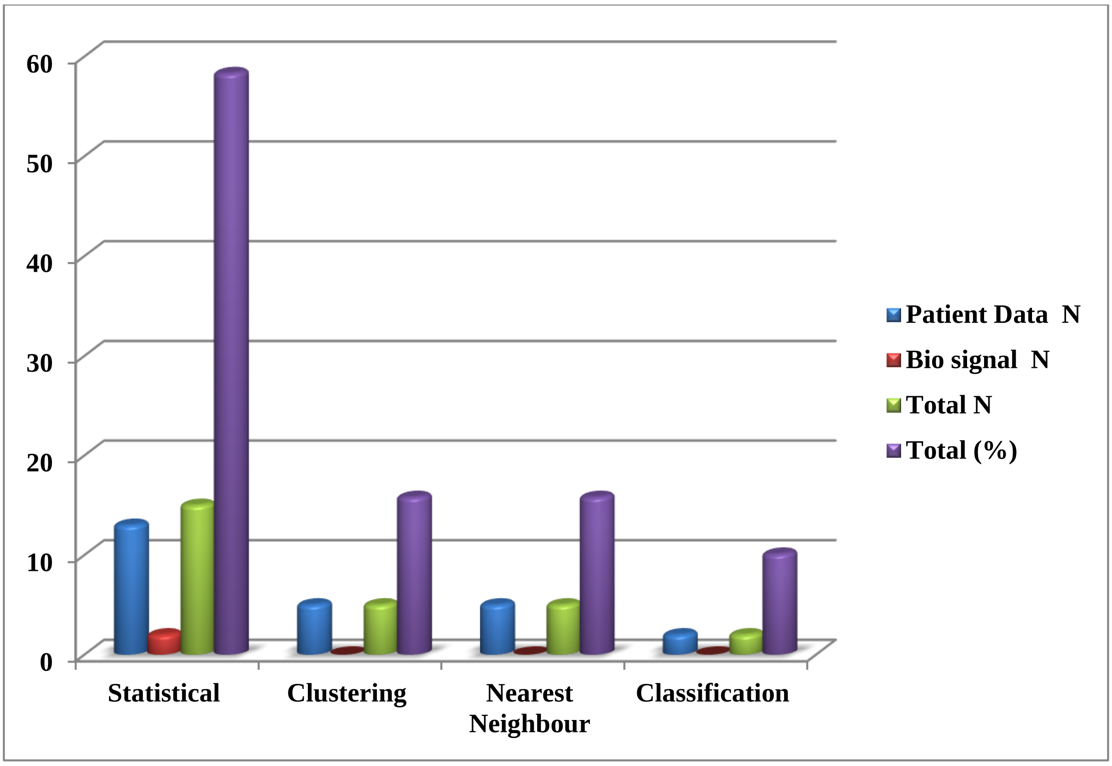


A comparison table between the methodology mentioned and the data types utilised in the study, "healthcare and quality indicator," which exclusively takes into account administrative medical data, is shown in Table 4.

Table: 4. Administrative Medical Data

	Data Type			
	Patient Data	Bio signal	Total	
Techniques	**N**	**N**	**N**	**(%)**
Statistical	13	2	15	58.3
Clustering	5	0	5	15.8
Nearest Neighbour	5	0	5	15.8
Classification	2	0	2	10.1
Total	25	2	27	100.0

Figure: 5. Chart of Administrative Medical Data



That's what table 4 shows "patient information" (91.6%) and "bio signal" (8.4%) make up most of the information used to break down administrative medical information, which is the main subject of this paper.

This study shows that when only administrative medical data are considered, clustering (58%), closest neighbour (15%), and other statistical techniques are used most frequently. In any event, while just taking administrative medical information into thought, the methods to a great extent use information, for example, "patient information." The two medical specialties that are most frequently employed are oncology and genetics (I. G. Lee, 2020). In the fields of medicine and public health, outlier detection frequently uses patient records. Outliers in the data can occur for a number of reasons, including aberrant patient conditions, apparatus problems, and recording errors. As a result, outlier detection is a very important issue in this field and calls for a high level of accuracy.

6. Conclusion

In this study, we find that statistical methods are frequently employed in administrative medical data, while closest neighbour and clustering data mining techniques are still rarely used in this context. In addition to the widely used statistics methods, there is a substantial business for developing methods based on clustering and nearest neighbour to find outliers in medical data management. The hospital's healthcare quality can be understood from the standpoint of data outliers using outlier identification based on these data. Basic information from this study might be useful for later research (R. Castro-Zunti, 2020).

Further research is warranted on the suggested outlier detection technique for identifying intricate relationships that affect medical outcomes by fusing data mining with patient records. By combining factor analysis with principal component extraction, the entire sets of interdependent relationships are analysed, and the quantity of the medical data is decreased from ten variables to three components. The discovery of three parameters for further investigation suggests that the application of data mining techniques to the detection of outliers in medical data is more effective and produces reliable extraction results.

The proposed model's outlier indexes have practical value, but there are still some flaws and room for improvement, particularly in the following areas:

- The development of medical big data has been severely hampered by the lack of interoperability between data mining knowledge and medical knowledge.
- The evaluation results are somewhat impacted by the relatively high proportion of missing fields, and the pre-processing step of the data is slowed down.
- The evaluation model's accuracy and usability will be improved by incorporating more data in future work. These data may include information on patient and employee surveys, hospital systems and processes, and information on medical safety and quality.

7. References

[1] A. Lal, Y. Pinevich, O. Gajic, V. Herasevich and B. Pickering, “Articial intelligence and computer simulation models in critical illness,” World Journal of Critical Care Medicine, vol. 9, no. 2, pp. 13–19,11204

[2] Banerjee S, Zare RN, Tibshirani RJ, Kunder CA, Nolley R, Fan R, Brooks JD, Sonn GA. Diagnosis of prostate cancer by desorption electrospray ionization mass spectrometric imaging of small metabolites and lipids. Proc Natl Acad Sci U S A. 2021;114(13):3334–9.

[3] Beam A, Kohane I. Big Data and Machine Learning in Health Care. J Am Med Assoc. 2021;319(13):1317–8.

[4] Brett Lantz, Machine Learning with R, Packt Publishing Limited, 2020. ISBN - 978-1782162148.

[5] C. Huang et al., “Sample imbalance disease classification model based on association rule feature selection,” Pattern Recognition Letters, vol. 133, pp. 280–286, May 2020, doi: 10.1016/j.patrec.2020.03.016.

[6] D. Hendrycks, M. Mazeika, and T. Dietterich, ‘‘Deep anomaly detection with outlier exposure,’’ 2021, arXiv:1812.04606

[7] Domingues, R.; Filippone, M.; Michiardi, P.; Zouaoui, J. A comparative evaluation of outlier detection algorithms: Experiments and analyses. Pattern Recognit. 2020, 74, 406–421.

[8] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2020;542(7639):115–8.

[9] Gibbons C, Richards S, Valderas JM, Campbell J. Supervised Machine Learning Algorithms Can Classify Open-Text Feedback of Doctor Performance With Human-Level Accuracy,. J Med Internet Res. 2021;19(3): 65.

[10] I. G. Lee, Q. Zhang, S. W. Yoon, and D. Won, “A mixed integer linear programming support vector machine for cost-effective feature selection,” Knowledge-Based Systems, vol. 203, p. 106145, Sep. 2020, doi: 10.1016/j.knosys.2020.106145.

[11] J. Hu, Y. Li, W. Gao, and P. Zhang, “Robust multi-label feature selection with dual-graph regularization,” Knowledge-Based Systems, vol. 203, p. 106126, Sep. 2020, doi: 10.1016/j.knosys.2020.106126.

[12] Jyothi, P.N.; Lakshmi, D.R.; Rao, K.V.R. A supervised approach for detection of outliers in healthcare claims data. J. Eng. Sci. Technol. Rev. 2020, 13, 204–213.

[13] Koufakou, A., and Georgiopoulos, M. (2020). A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. Data Min Knowl Discov 20, 259-289.

[14] L. Shinners, C. Aggar, S. Grace and S. Smith, “Exploring healthcare professionals’ understanding and experiences of articial intelligence technology use in the delivery of healthcare: An integrative review,” Health Informatics Journal, vol. 26, no. 2, pp. 1225–1211205

[15] R. Castro-Zunti, E. H. Park, Y. Choi, G. Y. Jin, and S. Ko, “Early Detection of Ankylosing Spondylitis using Texture Features and Statistical Machine Learning, and Deep Learning, With Some Patient Age Analysis,” Computerized Medical Imaging and Graphics, p. 101718, Apr. 2020, doi: 10.1016/j.compmedimag.2020.101718.

[16] R.Palaniappan, K.Sunderaj, S. Sundaraj, “A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals”, BMC Bioinformatics, 15.1, pp. 1-8, 2020.

[17] S. Das, M. R. Islam, N. K. Jayakodi, and J. R. Doppa, ‘‘Active anomaly detection via ensembles: Insights, algorithms, and interpretability,’’ 2021. arXiv:1901.08930.

[18] S. M. S. Shah, F. A. Shah, S. A. Hussain, and S. Batool, “Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methods,” Computers & Electrical Engineering, vol. 84, p. 106628, Jun. 2020, doi: 10.1016/j.compeleceng.2020.106628.

[19] S.A. Medjahed, T. A. Saadi, and A. Benyettou, "Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules." International Journal of Computer Applications Vol. 62 No. 1 , 2020.

[20] T. Davenport and R. Kalakota, “The potential for articial intelligence in healthcare,” Future Healthcare Journal, vol. 6, no. 2, pp. 94–102, 2021.

[21] W. Shu, W. Qian, and Y. Xie, “Incremental feature selection for dynamic hybrid data using neighborhood rough set,” Knowledge-Based Systems, vol. 194, p. 105516, Apr. 2020, doi: 10.1016/j.knosys.2020.105516.

[22] X. Qin, L. Cao, E. A. Rundensteiner, and S. Madden, ‘‘Scalable kernel density estimation-based local outlier detection over large data streams,’’ in Proc. EDBT, 2021, pp. 421–432.

[23] 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.

[24] 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

[25] 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

[26] 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

[27] Y. A. Qadri, A. Nauman, Y. Bin Zikria, A. V. Vasilakos and S. W. Kim, “The future of healthcare internet of things: A survey of emerging technologies,” IEEE Communications, Surveys and Tutorials, vol. 22, no. 2, pp. 1121–1167, 2020

[28] Y. Zhao, Z. Nasrullah, and Z. Li, ‘‘Pyod: A python toolbox for scalable outlier detection,’’ 2021, arXiv:1901.01588.

[29] Y. Zhao, Z. Nasrullah, M. K. Hryniewicki, and Z. Li, ‘‘LSCP: Locally selective combination in parallel outlier ensembles,’’ in Proc. SIAM Int. Conf. Data Mining, May 2021, pp. 585–593.