

Research Article

Effective of Smart Mathematical Model by Machine Learning Classifier on Big Data in Healthcare Fast Response

Mahmoud Ahmad Al-Khasawneh ¹, Amal Bukhari ², and Ahmad M. Khasawneh³

¹School of Information Technology, Skyline University College, University City Sharjah, 1797 Sharjah, UAE

²College of Computer Science and Engineering, University of Jeddah, Saudi Arabia

³Department of Cybersecurity, Amman Arab University, Amman 11953, Jordan

Correspondence should be addressed to Mahmoud Ahmad Al-Khasawneh; mahmoud@outlook.my

Received 11 December 2021; Revised 2 February 2022; Accepted 7 February 2022; Published 23 February 2022

Academic Editor: Deepika Koundal

Copyright © 2022 Mahmoud Ahmad Al-Khasawneh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the past few years, big data related to healthcare has become more important, due to the abundance of data, the increasing cost of healthcare, and the privacy of healthcare. Create, analyze, and process large and complex data that cannot be processed by traditional methods. The proposed method is based on classifying data into several classes using the data weight derived from the features extracted from the big data. Three important criteria were used to evaluate the study as well as to benchmark the current study with previous studies using a standard dataset.

1. Introduction

One of the important areas in artificial intelligence is machine learning. The discovery of the race is the main goal of machine learning, in addition to making smart decisions. There are a lot of algorithms for machine learning, and they are mainly classified into two types and supervised, and the second category is unsupervised, and there is a class between them called semi-supervised [1]. When there is big data, machine learning will be expanded by algorithms. Some researchers classify machine learning according to the categorized outputs such as regression, clustering, and density estimation, which mostly depend on teaching the decision tree, teaching the rules of association and artificial neural networks [2, 3].

Supervised algorithms include Naïve Bayes, boost algorithm, and support vector machines (SVM) that can be learned. Unsupervised learning takes undefined data and categorizes it based on features extracted and compared between them [4].

Big data is used in machine learning, which is a huge volume of structured or unstructured data. In which traditional methods are unable to process or by using traditional database methods, big data plays an active role in scientific

discoveries that depend on creating values, as well as massive parallel processing, distributed file systems, cloud computing, and the use of big data technologies such as Hadoop and server database SQL [5].

Big data originally shows the size, speed, and types of data, in addition to the time of production of that data by health care providers, and includes information related to patient care, and the magnitude comes from the abundance of its production and analysis. Demographics play the role of big data, diagnoses, medical treatments, drugs, lab results, and X-ray images. In the development taking place in the electronic renaissance, sensors, and flow tools, health care data has become of interest, despite the large amount of data it contains. Healthcare data is very useful in drug discovery, medical applications, disease prediction, and so on. Big health data plays an effective role in health care, scientific research, social networks, industry, and public administration. As for big data, it is mainly categorized in the form of 5Vs as follows [6, 7]:

- (i) Volume: large volume refers to big data. Big data contains a huge amount of information such as high-resolution videos, images, and text, in a

healthcare database; the data must be processed with traditional enhanced techniques for processing images, personal information, and readings from biometric sensors, which increases the complexity of the data due to the cumulative processing.

- (ii) **Velocity:** velocity represents the rate of creation of large stored data. The growth rate of paper data in social networking sites and public health is increasing, and because of the diversity as well, the processing speed should be concerned with X-ray films and text data.
- (iii) **Variety:** big data includes a variety of data that undoubtedly requires a special type of processing, such as formatted data, Excel, or CSV data that is stored in text files, such as doctor's prescriptions, clinical data, and office records.
- (iv) **Veracity:** it is the validity of the data that has no real errors, which is the ability to understand health care data and its quality in addition to verifying the testimony of information such as treatment, prescription, results, and procedures.
- (v) **Value:** big data takes into account the actual value of the generated data. Cost changes the importance of the value of data. Therefore, in health care, data and its cost are an actor in the system. High value to patients is the primary goal in health care.

2. Related Work

When thinking about machine learning, it means solving problems related to nontraditional ways. As for the data that is unusual, it is the big data that has properties that prevent it from being processed in traditional ways [8]. Big data needs a difficult and not natural prediction process, especially when we encounter abnormal data that needs to take special measures. The production of big data needs continuity of processing due to a small change in the data series that gives wrong results in the near future [9]. Academics are keen to keep track of big data and the technologies that deal with it, especially artificial intelligence. Machine learning includes working methods such as decision tree, SVM, Naïve classifier, industrial networking, clustering, and genetic computing [10, 11], all with one goal, which is to get future decisions and predictions, and with the goal of smart decisions for a machine learning strategy [12].

Big data analysis and processing are very important at the present time due to the development in information technology and the Internet. It provides in an equal way the use of computational and measurable data by machine learning of structured and unstructured data [13]. The problems with big data are not many if innovation and skill are combined. Big data may be in the biometric indicators in the google scholar through information such as the author's name, university, nationality, language, and number of references, and through the analyses of the above information, complex maps of the progress of the indexed process can

be created [14]. In the world of data, there is an increase in the rate of production nowadays, so the term big data was introduced. Therefore, analyzing such data has become one of the necessities that scientists have been finding compatible algorithms with [15]. Here, we will take a historical overview about big data and its relationship with machine learning and algorithms that were presented in previous studies as in Figure 1.

Big data analytics are associated with the work of large organizations and the areas of innovation and competitive production. We can define them as techniques deployed to reveal the hidden patterns of understanding the work contexts through the output. Each researcher divides machine learning according to how it works and how closely the topic relates to the main idea. Classification in this research focuses on machine learning and its classifications related to big data. Basically, classifications consist of many algorithms, but the most important of them are CVM, KNN, and ELM [16]. All are considered features extracted from the data in advance to be indexed based on certain criteria in the algorithm. The second and important category is the aggregation, which consists of several algorithms, but the main ones that work with big data are the *K*-Mean and DPSCAN algorithm [17], which depends on the relationship of the data with each other and the extent of the impact of one on the other. As for the last type, the Volitional Algorithm, it is widely used with data that has an interactive nature, meaning that the big data is continuously updated over time. Such algorithms are not useful in the medical field, which deals with data of great importance to human lives, especially those that rely on unanimous decisions due to the seriousness of the situation. Among these algorithms are GA, ACO, and PSO [18]. As conclusion many algorithms suggested in literature, each has advantage and disadvantage so that we try to adapt algorithm with minimum errors.

3. Big Data and Machine Learning

Big data applications are important in our daily lives. Among these applications is health care for patients that need to manage their data, especially with the spread of epidemics at the present time.

3.1. Big Data in Healthcare. According to IDC reports, the growth of big data in healthcare is expected to accelerate compared to other areas such as industries, financial services, and media. Health care data is expected to reach 36% until 2025, according to projections Compound Annual Growth Rate (CAGR) [19]. The global big data market is expected to reach \$34.27 billion in healthcare in 2022 if the current annual growth rate continues at 22%, which will reach the health sector with a budget of \$68 billion [20], so focusing on this area will save billions of dollars which is the main objective in this research. Big data in health care take the data from different resources as shown in Figure 2

Some resources provide data online that some time increased exponentially by the time, and most of the data comes from the document of hospital or patient himself.

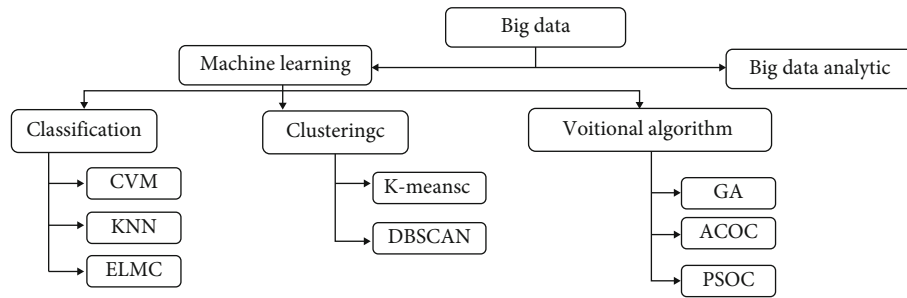


FIGURE 1: Structure of big data related to machine learning in existing studies.

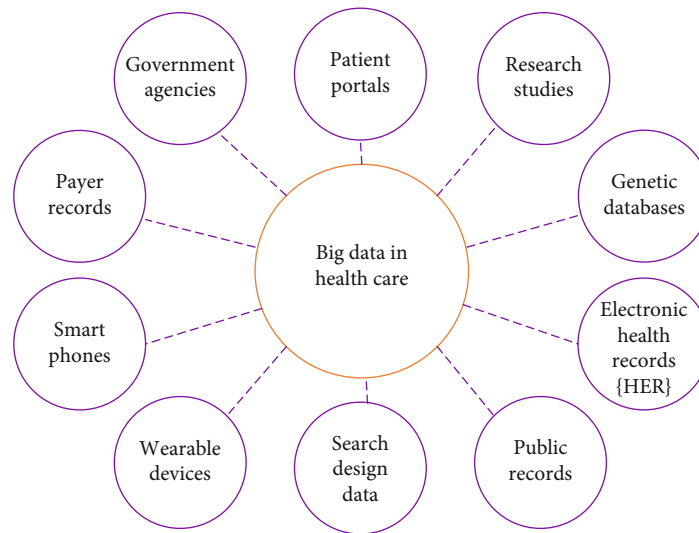


FIGURE 2: Big data health care resources.

Big data in the field of health care comes from several sources, starting from electronic records, search engines, and sensors, thus obtaining endless data that needs processing and analysis to know its use in the correct way. From this point of view, the usefulness of medical big data is at all levels, from the patient to the hospital and to society in general.

3.2. Machine Learning. Machine learning is considered a part or branch of artificial intelligence, which refers to the ability of information technology systems to find the best solutions for multiple patterns in big data. Machine learning depends on deducing algorithms and solutions in several courses and choosing the best among them. Machine learning algorithms rely on prior data extracted from master data called features [21]. Synthetic knowledge is derived from experiences in machine learning. For machine learning, statistical methods and mathematical processes are used in order to exploit computer resources as a whole.

Specifically, the computer program learns from experience (E) while executing some tasks (T) in order to measure performance (P), all of this to improve performance and get better results. The summary of the work is that the computer program can learn to improve perfor-

mance, and therefore, it can be measured for some tasks through experience. The number of resources within big data needs some processes to get performance p_{+1} which is smaller than P_i such as $\{P_i, P_{i+1}\} \subset P$ to get experience of E in the context of $\{e_i, e_{i+1}\} \subset E$, respectively. Then, the machine learning model is aimed at improving each performance till exceeding all iterations in the tasks. ML has basically four paradigms as shown in Figure 3 with brief summarization.

- (i) Unsupervised model is an experiment that contains many features of big data in order to identify the useful properties of group structuring, and in the field of machine learning, knowing the distribution possibilities is of great importance in order to choose the right group with high impact, and clustering is an example of unsupervised learning models [22]
- (ii) Supervised model is the experience of a set of data that contains features, but these features are related to a specific label or goal. The label is either binary (0 and 1) and indexed on the basis of their proximity to the correct output, or they are pretrained, and thus, new data is checked based on prior training [23]

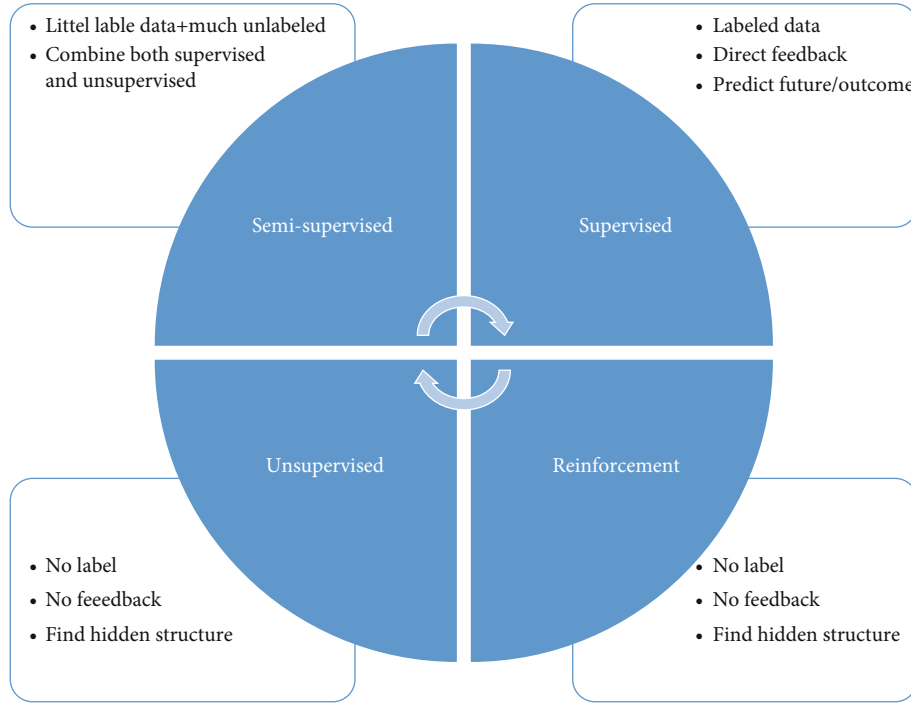


FIGURE 3: Type of ML paradigms.

- (iii) Semisupervised model is a homogeneous mixture of supervised and unsupervised models to obtain an individual accurate result for the specific tasks. Sometimes, it is called transformative education or inductive education [24]
- (iv) Reinforcement depends on interaction with the environment and the work itself by feeding back learning and acquiring new real-time experiences [25]

On the basis of these models, the linkage is revealed in the privacy of each of them. In general, the supervised and unsupervised models are trained on tasks (T) in the data that produce temporary jobs during implementation. Improving performance (P) is the main goal in some tasks. Training on real-time data that comes directly from the user is somewhat sensitive as in social media and contains privacy. Therefore, machine learning detects future privacy intentions, which is the goal of this study.

3.3. Challenging of Machine Learning in Big Data. There are many challenges related to the use of machine learning techniques in big data, including the following: (1) flexibility of computational design and the ability to develop machine learning through it; (2) through machine learning algorithms, being able to understand the characteristics and importance of data before applying the algorithm; and (3) ability to learn and build architecture and adapt to increasing volumes of labeled data.

Among the problems that we face when using machine learning with big data, which sometimes makes it inappropriate [26], are as follows: (1) The first one is not adapting

to the data, which means that some algorithms are suitable for certain data and not suitable for other data. (2) Training, which depends on a machine learning algorithm, uses dataset labeled for classification, and dataset that does not have label is inaccurately classified. (3) A particular algorithm may be good for certain data and bad for other data. One of the most important challenges facing machine learning is to expand the scope of big data to appropriate dimensions [8, 27, 28]. There are challenges related to the speed of data processing, as machine learning algorithms need time to train before starting the actual processing. Therefore, machine learning algorithms are designed based on the ability to analyze and the type of data to be processed.

To increase the efficiency of the system's work, the data is redesigned using SMART PLS that first classifies the data before starting to implement the machine learning algorithm.

4. Materials and Methods

Machine learning depends on its work on two methods; the first and the most important method is the classification process. Big data is initially initialized and indexed into aggregates and by master key. Big data analysis is preliminary, such as $s(\text{bigdata analytic}) < \min(s(\text{bigdata}), s(\text{analytic})) = s(v_1 \vee v_2)$, where s is a group of data, \vee is or operation, and v is considered the value sequence. Let $v_1, v_2, v_3 \in V$ where V is the big data value; then $s(v_1) = F(v_1), s(v_2) = F(v_2)$ and $s(v_3) = F(v_3)$, so the main function will be like

$$s(v_1 \vee v_2) = s(v_1) \cap s(v_2) = F(v_1) \cap F(v_2). \quad (1)$$

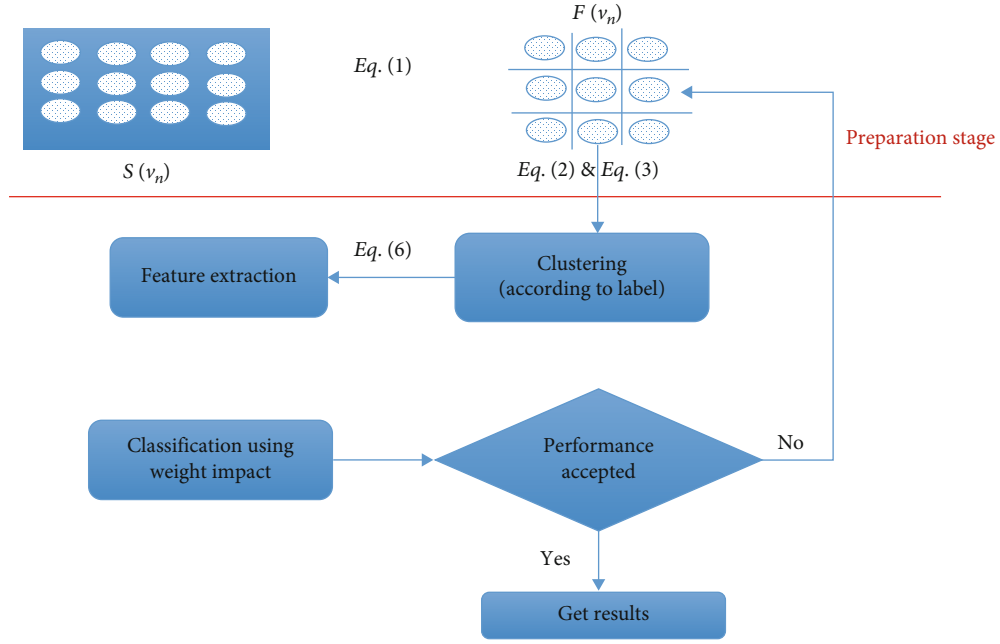


FIGURE 4: Big data process within ML.

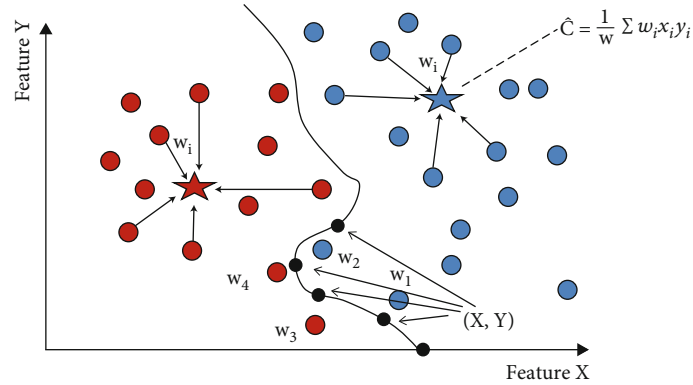


FIGURE 5: Classification underweight condition.

By Equation (1), we can cluster the big data into sub-groups with its function and using SMART PLS program to calculate the weights that can help at the next step of analytics. Statistical program was used to arrange the data into appropriate sections for better preparation. Due to $s(v_i), i = \{1, 2, 3, \dots, n, \dots\}$, then the whole group of big data is illustrated in

$$s(v_1 \vee v_2 \vee v_3 \dots \vee v_n) = \prod_1^n F(v_i) = F(v_1) \cap F(v_2) \cap \dots \cap F(v_n). \quad (2)$$

This make relation of the entire big data within a subgroup achieve n groups clustered in certain condition. And because of data availability that will be still updated or online produced, the limit of n will change to infinity

$$s(v_1 \vee v_2 \vee v_3 \dots \vee v_n \dots) = \prod_1^\infty F(v_i). \quad (3)$$

And with respect to the operation *or* (\vee) in case of updated data, but when produced, new data belong to original big data operation *or* will change to *and* (\wedge) as

$$s(v_1 \wedge v_2 \wedge v_3 \dots \wedge v_n) = \prod_1^n F(v_i) = F(v_1) \cup F(v_2) \cup \dots \cup F(v_n), \quad (4)$$

$$s(v_1 \wedge v_2 \wedge v_3 \dots \wedge v_n \dots) = \prod_1^\infty F(v_i). \quad (5)$$

This process occurs in preparation data or initialization stage after the big data will be ready to extract features as shown in Figure 4.

TABLE 1: Attribute within UCI ML dataset.

encounter_id	patient_nbr	Race
Gender	age	Weight
admission type id	discharge disposition id	admission_source_id
time_in_hospital	payer_code	medical_specialty
numb_lab_procedures	num_procedures	num_medications
number_outpatient	number_emergency	number_inpatient
diag_1	diag 2	diag 3
number_diagnoses	max_glu_serum	AlCresult
Metformin	repaglinide	Nateglinide
Chlorpropamide	glimepiride	Acetohexamide
Glipizide	glyburide	Tolbutamide
Pioglitazone	rosiglitazone	Acarbose
Migitol	troglitazone	Tolazamide
Examide	citoglipton	Insulin
glyburide-metformin	glipizide-metformin	glimepiride-pioglitazone
metformin-rosiglitazone diabetesMed	metformin-pioglitazone readmitted	Change

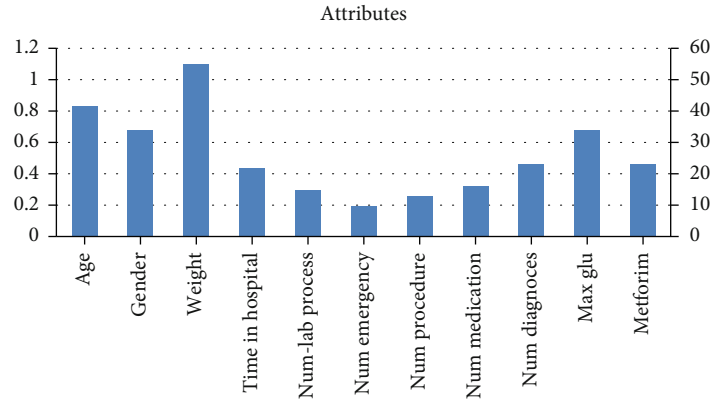


FIGURE 6: Database attributes after reducing.

Patient behavior was predicted by using linear predictor for j data point such as $j = 1, 2, 3, \dots, n$.

$$f(j) = a_0 + a_1x_{j1} + a_2x_{j2} + \dots + a_mx_{jm}. \quad (6)$$

In the above equation, x_j is the feature vector for $j = 1, 2, \dots, m$ of m -th interactive variable of data point value j , and $a_0, a_1, a_2, \dots, a_m$ are the coefficients of the results that presentative relation between the data since the coefficient is accumulative into the vector. Feature extraction based on j -th cluster of M features $\langle x_1, x_2, \dots, x_n \rangle$ such as n refer to cluster number. Weight that controls the features $\langle w_1, w_2, \dots, w_m \rangle$ was also used to calculate \bar{x} of mean value belonging to feature x

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (7)$$

From above equation, we can get the standard deviation σ shown in

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i^j - \bar{x}^j)^2}{N-1}}. \quad (8)$$

Standard deviation was related to new features of certain weight and expressed as

$$\bar{x}_i = \frac{x_i}{\sigma}. \quad (9)$$

Then, weight can be calculated as

$$w_i = \frac{1}{\sigma}. \quad (10)$$

Figure 5 shows behavior of classifier when weight value is variable including some conditions.

TABLE 2: Evaluation criteria.

Parameters	Proposed weighted classification	DT outlier	DTI	RBF
Precision	0.99	0.99	0.89	0.82
Probability of detection	0.76	0.7	0.65	0.87
<i>F</i> -score	0.83	0.82	0.73	0.84

TABLE 3: Confusion matrix.

Predicted class		Yes	No	Total
Actual class	Yes	TP	FN	<i>P</i>
	No	FP	TN	<i>N</i>
	Total	<i>P'</i>	<i>N'</i>	<i>P + N</i>

Big data will train (x_i, y_i) for N iteration to i , with $x_i \in R$ and $y_i \in \{-1, 1\}$, of learning classifier Y such that

$$Y = \begin{cases} +1, & \text{if } x_i \in \text{class } A, \\ -1, & \text{if } x_i \in \text{class } B. \end{cases} \quad (11)$$

The classification of the data is in a sequential form, and this sequence can be rearranged in the event that there are cutoffs, such as updating the data or adding supplementary data to the previous data. In Results and Discussion, we will discuss training and data testing and the method of implementing the classifier for big data.

5. Results and Discussion

The proposed algorithm was implemented using MATLAB language and Windows 10 operating system on standard dataset called UCI ML [29] and with a 50 of attributes as in Table 1

The dataset consisted of 50 attributes with 101,767 patient records. The processing was calculated and divided into two parts: the first section is training, with 80%, with the data that contains a label, and the second section is testing, with 20%. The first big dataset reduced into 20123 records when applying reduce procedure to be the best fit dataset even though some attributes also reduced for parallel execution [30, 31], after reducing some attributes in dataset highlighted such as gender, age, medical specialty, time in hospital, and weight, as shown in Figure 6.

We check the validity and reliability of the database attributes within the algorithms to find the relation among the data during processing to take effect of certain attributes. Such as $\{weight, age\} \rightarrow \{num\ emergency\}$. Classification process is done by multiclass technique to find three main evaluation factors such as precision, probability of detection, and *F*-score. And for benchmark results, then, we used standard dataset and each algorithm contains different result values, so efficiency of the proposed method is illustrated in the Table 2.

Better classification rate achieved due to reducing data was great, and the features extracted were useful, so predictions are classified as two cases correctly and incorrectly; many conditions were applied during creating confusion matrix such as probability of detection (sensitivity), accuracy, precision, true negative rate (specificity), and *F*-score [32]. The true positive detection occurs during emergency classification; then accuracy is formulated as ratio in confusion matrix. Equations below refers to criteria that can be detected by the proposed algorithm.

$$\begin{aligned} \text{Accuracy} &= \frac{(TP + TN)}{(TP + FP + FN + TN)}, \\ \text{Probability of detection} &= \frac{TP}{(TP + FN)}, \\ \text{True Negative} &= \frac{TN}{FP + TN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \end{aligned} \quad (12)$$

where TP represents the count of emergency case classified as emergent patient, TN refers to normal classified, FP refers to the number of emergent case, and FN refers to the number of normal classified. Most of results depend on confusion matrix and its condition as shown in Table 3.

In Table 3, TP refers to True Position, TN refers to True Negative, FP refers to False Position, and FN refers to False Negative. In addition, these criterion measurement classifiers can also be doing additional aspects such as robustness, speed, scalability, and interpretability.

6. Conclusion

The evaluation of big data is often useful for health care, which has received high interest nowadays due to the spread of viruses and epidemics in general. The data coming from the field of health care in general is of two types, either updateable or in the form of large data. Classifying such data requires a complex computational process and the development of a classifier adaptable to that data. In this study, a classifier was improved based on the weight of the data, and this weight is derived from the extracted features in order to classify the medical data and according to the importance. At first, the data is reduced in order to be processed as quickly as possible, and then it is classified on the basis of several types. The main attributes such as age, medical prescription, number of visits to the doctor, hospitalization periods, medical analyses, and others are considered complementary relationships for classification. The study was evaluated by three criteria: precision, probability of detection, and *F*-score used for benchmarking on standard dataset; meanwhile, other evaluations such as accuracy and true negative are also considered in this study.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] L. Gutiérrez-Gómez, F. Petry, and D. Khadraoui, "A comparison framework of machine learning algorithms for mixed-type variables datasets: a case study on tire-performances prediction," *IEEE Access*, vol. 8, pp. 214902–214914, 2020.
- [2] J. Z. Kim, Z. Lu, E. Nozari, G. J. Pappas, and D. S. Bassett, "Teaching recurrent neural networks to infer global temporal structure from local examples," *Nature Machine Intelligence*, vol. 3, no. 4, pp. 316–323, 2021.
- [3] S. A. Ali Shah, I. Uddin, F. Aziz, S. Ahmad, M. A. Al-Khasawneh, and M. Sharaf, "An enhanced deep neural network for predicting workplace absenteeism," *Complexity*, vol. 2020, 12 pages, 2020.
- [4] G. Sulong and A. Mohammedali, "Recognition of human activities from still image using novel classifier," *Journal of Theoretical and Applied Information Technology*, vol. 71, no. 1, 2015.
- [5] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander, and M. S. H. Sunny, "Application of big data and machine learning in smart grid, and associated security concerns: a review," *IEEE Access*, vol. 7, pp. 13960–13988, 2019.
- [6] S. Khan, S. Nazir, I. García-Magariño, and A. Hussain, "Deep learning-based urban big data fusion in smart cities: towards traffic monitoring and flow-preserving fusion," *Computers and Electrical Engineering*, vol. 89, article 106906, 2021.
- [7] B. Rawat and J. K. Samriya, "A study on challenges of big data and their approaches in present environment," *Proceedings of Integrated Intelligence Enable Networks and Computing*, pp. 483–495, Springer, Singapore, 2021.
- [8] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-AI integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.
- [9] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 989–1037, 2020.
- [10] B. T. Atiyha, S. Aljabbar, A. Ali, and A. Jaber, "An improved cost estimation for unit commitment using back propagation algorithm," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 15, no. 2, pp. 243–248, 2019.
- [11] F. Padillo, J. M. Luna, and S. Ventura, "A grammar-guided genetic programming algorithm for associative classification in big data," *Cognitive Computation*, vol. 11, no. 3, pp. 331–346, 2019.
- [12] J. Ma, Y. Ding, J. C. P. Cheng et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques," *Journal of Cleaner Production*, vol. 244, article 118955, 2020.
- [13] K. A. Brown, S. Brittan, N. Maccaferri, D. Jariwala, and U. Celano, "Machine learning in nanoscience: big data at small scales," *Nano Letters*, vol. 20, no. 1, pp. 2–10, 2020.
- [14] J. B. Awotunde, R. G. Jimoh, I. D. Oladipo, M. Abdulraheem, T. B. Jimoh, and G. J. Ajamu, "Big data and data analytics for an enhanced COVID-19 epidemic management," in *Artificial Intelligence for COVID-19*, pp. 11–29, Springer, Cham, 2021.
- [15] L. Zhang and N. Li, "Material analysis and big data monitoring of sports training equipment based on machine learning algorithm," *Neural Computing and Applications*, vol. 34, no. 4, pp. 2749–2763, 2022.
- [16] G. Hernández, E. Zamora, H. Sossa, G. Téllez, and F. Furlán, "Hybrid neural networks for big data classification," *Neurocomputing*, vol. 390, pp. 327–340, 2020.
- [17] M. Nayak and V. Sharma, *New Modified K-Mean Cluster Algorithm (NMKMCA) in Online Prediction of Product Price in Bigdata*, NOVY MIR Research Journal, 2021.
- [18] A. K. Dubey, A. Kumar, and R. Agrawal, "An efficient ACO-PSO-based framework for data classification and preprocessing in big data," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 909–922, 2021.
- [19] B. N. Subudhi, D. K. Rout, and A. Ghosh, "Big data analytics for video surveillance," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 26129–26162, 2019.
- [20] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, pp. 1–25, 2019.
- [21] G. Sulong and A. Mohammedali, "Human activities recognition via features extraction from skeleton," *Journal of Theoretical and Applied Information Technology*, vol. 68, no. 3, 2014.
- [22] A. Mohammadnazar, R. Arvin, and A. J. Khattak, "Classifying travelers' driving style using basic safety messages generated by connected vehicles: application of unsupervised machine learning," *Transportation Research Part C: Emerging Technologies*, vol. 122, article 102917, 2021.
- [23] M. S. Rehan, F. Rustam, S. Ullah, S. Hussain, A. Mehmood, and G. S. Choi, "Employees reviews classification and evaluation (ERCE) model using supervised machine learning approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, pp. 1–18, 2021.
- [24] M. Aamir and S. M. A. Zaidi, "Clustering based semi-supervised machine learning for DDoS attack classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 4, pp. 436–446, 2021.
- [25] M. K. Eckstein, L. Wilbrecht, and A. G. E. Collins, "What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience," *Current Opinion in Behavioral Sciences*, vol. 41, pp. 128–137, 2021.
- [26] M. Krishnan, "Against interpretability: a critical examination of the interpretability problem in machine learning," *Philosophy & Technology*, vol. 33, no. 3, pp. 487–502, 2020.
- [27] H. M. Verhelst, A. W. Stannat, and G. Mecacci, "Machine learning against terrorism: how big data collection and analysis influences the privacy-security dilemma," *Science and Engineering Ethics*, vol. 26, no. 6, pp. 2975–2984, 2020.
- [28] I. Kotenko, I. Saenko, and A. Branitskiy, "Machine learning and big data processing for cybersecurity data analysis," in *Data Science in Cybersecurity and Cyberthreat Intelligence*, pp. 61–85, Springer, Cham, 2020.

- [29] D. Dua and C. Graff, *Machine Learning Repository*, University of California, Irvine, School of Information and Computer Sciences, 2017, <http://archive.ics.uci.edu/ml>.
- [30] M. A. Saleh, S. H. Othman, A. Al-Dhaqm, and M. A. Al-Khasawneh, "Common investigation process model for Internet of Things forensics," in *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Cameron Highlands, Malaysia, 2021.
- [31] M. I. Uddin, S. A. Ali Shah, M. A. Al-Khasawneh, A. A. Alarood, and E. Alsolami, "Optimal policy learning for COVID-19 prevention using reinforcement learning," *Journal of Information Science*, vol. 165551520959798, p. 016555152095979, 2020.
- [32] I. Ahmad, S. A. A. Shah, and M. A. Al-Khasawneh, "Performance analysis of intrusion detection systems for smartphone security enhancements," in *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Cameron Highlands, Malaysia, 2021.