Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

A Survey on Health Care and Expert System

Dr. Asha Ambhaikar

Kalinga University, Raipur, Chhattisgarh, India
dsw@kalingauniversity.ac.in

Article Info
Page Number: 451-461
Publication Issue:
Vol. 72 No. 1 (2023)

Article History
Article Received: 15 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

Abstract
Since 2010, big data has been more prevalent in healthcare due to three primary factors, including the vast amount of data available, the rising expense of healthcare, and a focus on individualized health-care services. Data that is too large or complicated for traditional data processing methods to process is referred to as "Big Data" in healthcare. A few examples of large-scale data sources in healthcare are the Internet of Things (IoT) and Electronic Medical Records (EMRs/EHRs), which contain patient medical history and diagnose information, medication regimens and treatment plans for a patient's condition as well as allergy information and results from laboratory and test tests, genomic sequencing, medical imaging and other clinical data sources. A variety of machine learning algorithms were used to analyze healthcare data in this study, and the results were discussed. Determining how to deal with huge data as well as the apps that use it. Using machine learning techniques and the necessity to handle and utilize massive data from a fresh perspective is the focus of the paper.

Keywords: - Health Sector; Big data; Machine Learning (ML); artificial intelligence (AI).

I. INTRODUCTION

"Big data" refers to a collection of data generated by health care providers at different times in time. In addition to demographics, diagnoses, medical procedures and medications, vital signs and vaccines are included in this data as are laboratory findings and radiologic pictures. As medical data collecting increases, electronic health sources such as sensor devices, streaming machines, and high-throughput instruments are acquiring more data. According to Chandu Thota et al. [1,] this health sector big data is used for a variety of applications including diagnosis, drug development, precision medicine, disease prediction, and so on. Big data has become more important in a range of fields, including health sector, scientific research, industry, social networking, and government administration [1]. The following 5Vs can be used to categorize big data: Big data is certainly represented by the large volume. Traditional data processing platforms and methodologies must be enhanced in order to process massive amounts of data such as text, audio, video, and large-scale images. Personal information, radiological images, personal medical records, genomes, and biometric sensor readings, among other things, are gradually being included to a health sector database. The size and complexity of the database are greatly increased as a result of all of this information. Big data is defined by its velocity, or the rate at which data is

generated. The social media data explosion has transformed the data landscape and resulted in a wide range of data. The majority of health data is in the form of paper files, X-ray films, and scripts, and the rate of increase of such data is rapidly increasing. Variety: Big data is unquestionably represented by the diversity of data. Database, excel, and CSV, for example, are all forms of data formats that can be saved in a plain text file. There are three types of health data: structured, unstructured, and semi-structured. Clinical data is an example of structured data; nevertheless, data like doctor notes, paper prescriptions, office medical records, pictures, and radiological films are unstructured or semi-structured. Data veracity: This data veracity accurately reflects bug data. It denotes data readability rather than data quality. The veracity component in health sector data provides information certificates such as proper diagnosis and so on. Big data is actually represented by the value of data. When it comes to big data analytics, the benefits and costs of analyzing and acquiring big data are more essential. The rewards for all other actors in the health sector system should be determined by the provision of value for patients. The primary purpose of health sector delivery must be to provide great value to patients.

II. Big data in health sector has several applications

Big data applications open up new avenues for discovering new information and developing creative approaches for improving health-care quality. Public health, clinical operations, research and development, patient profile analysis, evidence-based medicine, remote monitoring, and other applications are among the most significant. The frameworks and storage systems for big data in health sector are depicted below.

Remote monitoring in the health sector industry is now possible thanks to Internet of Things (IoT)-enabled devices, which has the ability to keep patients safe and healthy while also empowering clinicians to provide superior treatment. As contacts with doctors have gotten easier and more efficient, it has also boosted patient participation and satisfaction. Furthermore, remote monitoring of a patient's health helps to shorten hospital stays and avoid re-admissions. IoT has a big impact on lowering health sector expenses and increasing treatment outcomes.

Epidemiology that uses digital approaches from data collecting to data analysis is known as digital epidemiology. Traditional epidemiological research, such as case records, case reports, ecological studies, cross-sectional studies, case-control studies, cohort studies, randomized controlled trials, and systematic reviews and meta-analysis, benefit from it. It also makes use of data sources that were initially collected and/or developed for medical and non-medical purposes.

Rather than a model of a specific infectious disease, FRED (a framework for reconstructing epidemiological dynamics) is an open-source framework for epidemic modeling. In FRED, geographic regions are used to represent each human as an agent. Age, sex, employment status, occupation, and household location, as well as membership in a set of social contact networks, are all sociodemographic features and everyday behaviors that each agent possesses. FRED uses this synthetic population data to model the disease outbreak.

A distributed NoSQL database is used to store a large amount of data. A relational schema is not followed by a NoSQL database. There are four types of NoSQL databases: key-value stores, column family database stores, document stores, and graph stores. The key-value store is used for tiny applications and saves data using key-value pairs. Second, as a collection of columns, the column family database holds large amounts of data in rows. Third, the document stores a lot of information about a document format and is utilized to store semi-structured data. Finally, the graph stores database has nodes with map and query relationships that have edges between them.

III. Big data in health sector and artificial intelligence

Machine Learning is a sub-discipline of artificial intelligence that refers to the ability of computer systems to find solutions to problems on their own by recognizing patterns in databases. Machine Learning allows IT systems to spot patterns and build appropriate solution concepts based on current algorithms and data sets. As a result, artificial intelligence generates artificial knowledge based on experience. To learn from data sets, statistical and mathematical methods are utilized in artificial intelligence. Symbolic approaches and sub-symbolic approaches are the two main systems. Sub-symbolic systems are artificial neuronal networks, whereas symbolic systems are propositional systems in which the knowledge content, i.e., the induced rules and examples, are openly expressed. These are based on the principle of the human brain, in which knowledge is implicitly represented. Large scale data, varied forms of data, high speed streaming data, unclear and incomplete data are all key difficulties in machine learning for big data [5].

IV. Supervised, unsupervised, and reinforcement learning are the three main types of machine learning.

To begin, supervised learning refers to a problem in which a model is used to learn a mapping between input examples and the target variable. As a result, it necessitates training with labeled data that includes both inputs and outputs.

Second, unsupervised learning refers to a set of challenges in which a model is used to characterize or extract correlations in data without the use of labeled training data or the presence of desired goals in the environment.

Third, reinforcement learning refers to a set of issues in which an agent must learn to operate in a given environment utilizing feedback. • Representation learning • Active learning • Deep learning • Transfer learning • Distributed and parallel learning • Kernel-based learning are some additional important learnings for handling big data problems.

Deep learning plays a significant role in the field of health sector among all of these. Deep learning is a sort of machine learning that solves issues that machine learning couldn't handle. Neural Networks are used in deep learning to boost computing work and offer correct results. Medical practitioners and researchers are using deep learning to uncover hidden opportunities in data and better serve the health sector business.

V. Deep learning for huge data in health sector

Deep neural networks (DNNs) are the state-of-the-art in machine learning and big data analytics, and they're employed in a wide range of applications, from defense and surveillance to human-computer interaction and question answering. DNN architecture comes in a variety of shapes and sizes, which can be divided into three categories. Feed-forward Neural Networks (FNN), Convolution Neural Networks (CNN), and Recurrent Neural Networks (RNN) are the three types [6]. Deep learning in health sector allows clinicians to precisely analyze any ailment and assist them treat it, resulting in improved medical judgments. Deep learning technology can be used in hospital management information systems to save costs, shorten hospital stays, control insurance fraud, predict changes in illness patterns, provide high-quality health sector, and improve medical resource allocation efficiency.

Several application examples are described in the following paragraphs based on the diverse types of biomedical information: biomedical pictures, biomedical time signals, and other biomedical data such as laboratory findings, genomics, and wearable devices.

Drug discovery Deep learning in health sector aids in the discovery and development of new treatments. As a result of the analysis of the patient's medical history, the most appropriate treatment is determined. Aside from that, patients' symptoms and testing are used to inform this technology.

Medical imaging In the diagnosis of disorders such as heart disease or brain tumor, medical imaging procedures such as MRI scans, CT scans, and ECG are used. As a result, deep learning aids clinicians in better diagnosing and treating patients.

Insurance fraud Medical insurance fraud claims are analyzed with deep learning. As a result of predictive analytics, it is possible to predict future fraud claims. The insurance business also benefits from deep learning by being able to send out discounts and offers to their target clients.

Genome In order to analyze a genome and provide patients a notion of the disease that could impact them, deep learning techniques are applied. Genomics and the insurance business are two areas where deep learning has a bright future ahead of them. Deep learning algorithms are used by Cells cope to let parents monitor their children's health in real time via a smart gadget, reducing the need for frequent doctor visits. Researchers believe that deep learning in health sector can be used to improve medical treatment for patients and professionals alike.

VI. Literature survey

Heart disease and sepsis prediction using machine learning

When the heart muscle is unable to pump enough blood, the body's needs for blood and oxygen are not met. Unfortunately, the heart can't keep up. The death rate from heart failure is comparable to or even higher than that from different malignancies. However, unexpected death has also been described as a regular cause of death. A patient's electronic health record (EHR) contains diagnostic information about the patient, physician information, and hospital departmental information. We can collect unstructured data in bulk from EHR time series for heart failure. These time-based EHRs can be analyzed and mined in order to establish correlations between diagnostic events and anticipate when a patient will be diagnosed. As a result of EHR data's structure and scope (including provider conduct, service utilization, treatment routes, and patient disease cases), as well as its uneven sample frequency, EHR data are extremely difficult to interpret and interpret.

It was proposed that SHFM (Seattle Heart Failure Model) be used in conjunction with EHR at Mayo Clinic in order to construct a risk prediction model based on machine learning techniques applied to routine clinical care data [12]. One of the most often used models for HF survival risk assessment is the Seattle Heart Failure Model (SHFM), which incorporates the influence of HF therapy on patient outcomes. Using data from 119,749 Mayo Clinic patients between 1993 and 2013, the researchers detected 5044 people with HF after applying some particular criteria and excluding the number of patients due to insufficient data. I HF survival prediction models constructed on EHRs are more accurate than SHFM, (ii) include co-morbidities improves the accuracy of the models, and (iii) there are potential interactions between diagnosis history, co-morbidities, and survival risk. As a consequence of utilizing a variety of machine learning methods, Logistic Regression and Random Forest produced the most accurate classifiers.

For example, Andy Schuetz and his colleagues developed RNN models that used gated recurrent units (GRUs) to find relationships between time-stamped events (e.g., disease diagnosis, medicine prescriptions, procedure orders, etc.) across a 12- to 18-month observation period for patients and controls [13]. Compared to many older methods, the RNN delivers a nonlinear development in model generalization as well as more scalability A health system's EHR provided 3884 incident HF cases and 28,903 controls between May 16, 2000 and May 23, 2013. The one-hot vector format, which is commonly used for NLP (natural language processing) activities, was used to describe clinical events in EHR data as computable event sequences. Comparing regularized Logistic Regression, Neural Network, Support Vector Machine and K-Nearest Neighbor classifier techniques, the model's performance metrics were compared. Several types of RNN models have been created using temporal sequenced data.

Cancer prediction using machine learning

For breast cancer patients with EHRs, Susan E. Clare et al. suggested a unique concept-based filter and prediction model for local recurrence detection [17]. Machine learning and Natural Language Processing (NLP) are employed to find these recurring patterns. Northwestern Memorial Hospital patients diagnosed with breast cancer between 2001 and 2015 were identified by (International Classification Diseases) In this study, ICD9 codes were

employed. On the basis of an analysis of 50 progress notes, partial sentences were taken out that indicated the local recurrence of breast cancer. They were then processed using MetaMap to provide a positive set of concepts called features. In order to map biomedical text to the UMLS (Unified Medical Language System) Metathesaurus, or, in other words, to identify Metathesaurus concepts alluded to in English text, the Lister Hill National Center for Biomedical Communications at the National Library of Medicine (NLM) developed MetaMap. To train a Support Vector Machine to detect local recurrences, these attributes are merged with the amount of pathology reports for each patient. Other baseline classifiers could not match this model's AUC.

For breast cancer prediction, Shruti Garg et al. used hyperparameters to create a machine learning model. As part of this research, six machine learning algorithms were used, including K-Nearest Neighbour, Logistic Regression, the Decision Tree (DT), the Random Forest (RF), the Support Vector Machine (SVM), and deep learning utilizing artificial neural networks (ANN). Adam Gradient Descent cost function was utilized for deep learning. Using Adam instead of the standard stochastic gradient descent approach, network weights can be iteratively updated based on training data. Adaptive gradient algorithm (AdaGrad) and root mean square propagation are combined in Adam learning (RMSProp). They were acquired from the Wisconsin Breast Cancer Dataset (WBCD), which already classifies malignant and benign breast cancers according to their severity. Using fine-needle aspiration (FNA) of the breast mass, 30 characteristics were calculated. When compared to other machine learning algorithms, Deep Learning achieves a better level of accuracy.

For breast cancer prediction, Alkhawaldeh et al. suggested a Multi-Layer Perceptron (MLP) model. On one hand, we looked at reducing the extracted feature size, while on the other, we looked at improving classification power. There are 569 instances and 32 attributes in the WDBC dataset. As the name suggests, Tuned MLP is based on shrinking the extracted feature. We employed a variety of search strategies and attribute assessors, and then voted on the most important features from the WDBC dataset, reducing the number of features from 31 to only four. By using Grid Search to determine the ideal hyperparameters of a model, Tuned MLP can make more "accurate" predictions than basic MLP.

Diabetes prediction using machine learning

Convolution neural network-based multimodal disease risk prediction (CNN-MDRP) algorithm has been developed by Kai Hwang et al. for big data [7]. Using a combination of structured and unstructured information, the disease risk model is created and its correctness is evaluated. It was more accurate than the CNN-based unimodal disease risk prediction algorithm (CNN-UDRP). EHR, medical images, and gene data were all obtained from a hospital in China. Patient department information is primarily constituted of unstructured and organized textual data. The CNN-MDRP algorithm's parameter is trained using the stochastic gradient descent algorithm, which is designed for big data applications. By employing big data analytics and the Hadoop platform, health sector providers may now obtain insight from clinical datasets and make informed decisions. People's requirements can be predicted with the help of data-driven services, which can be used to improve health sector management.

An EHR database that is large data was used by Ya Zhang and Tao Zheng to offer a semi-automated system based on machine learning. 15 local EHR systems in China were used to collect the data, which was automatically uploaded to the centralized repository every 24 hours. Based on a supervised learning algorithm, the framework was developed. In order to properly organise the raw EHR, feature engineering was needed. I extracted and created 16 features for use in the machine learning framework that I'm currently working on. Using the machine learning methods Random Forest, Logistic Regression, and Ada Boost, we were able to improve the accuracy of our predictions. To enhance recall while maintaining low false-positive rates, the algorithms also refine the filtering criteria.

An EHR-based machine-learning system was suggested by Ya Zhang et al. to identify type 2 diabetes. Three years of electronic health record data were used in this study. In Changning, Shanghai, the data was held in a central repository, which has been handled by the Changning District Health Bureau since 2008. On an hourly basis, electronic health record data from ten local EHR systems is automatically uploaded to the centralized repository. These include K-Nearest-Neighbors (KNN), Nave Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR). Also, the algorithm's identification performance was superior to that of the current state-of-the-art.

It was proposed by Dilip Singh Sisodia et al. These studies were undertaken in an attempt at developing an accurate diabetes prognostication model for patients with a high risk of developing diabetes. These algorithms included SVM, Naive Bayes and Decision Trees. All the data was gathered from the UCI machine learning repository's Pima Indians Diabetes Database (PIDD). Nave Bayes was the most accurate of the three machine learning algorithms tested.

Table 1. Precision/AUC of machine learning algorithms based on big data in healthsector.

Reference	Year	Machine learning techniques	Accuracy(%)/AUC
[20]	2020	Support Vector Machine (SVM)	86%
[18]	2020	Deep Learning-Artificial Neural Network (DL-ANN)	98%
[19]	2019	Multilayer Perceptron (MLP)	97%
[22]	2019	Elastic net	74%
[23]	2019	Neural Network (NN)	71%
[14]	2018	Long Short Term Memory (LSTM)	64%
[11]	2018	Naive Bayes (NB)	76.3%
[17]	2018	Support Vector Machine (SVM)	93%

VII. Machine learning models are evaluated using big data classification and measurements

Unstructured big data is classified using classification, a data mining process that aids in knowledge discovery and future ideas. As a result of classification, you can make innovative decisions. Two phases of classification exist. The first is the learning process where massive training data sets are provided and inquiry takes place, followed by the creation of rules and patterns that may be used to classify the data. As a result of this review of data sets and archives, the second phase begins to be executed. [24] A binary search tree (BST) was created and utilized in conjunction with a KNN to speed up the process of classifying large amounts of huge data. There is a method for sorting examples of the training data into a BST that relies on identifying the furthest pair of points (diameter). A pair of the BST's furthest points is found at each node. Then, all of the examples situated at that node are further sorted depending on their distances to these local furthest points It is then utilized to find a test example on the leaf and categorize it using the KNN classifier, using the BST that has been built. For big data categorization, Yi-Hung Huang et al. suggested an Improved Cat Swarm Optimization (ICSO) technique [25]. This ICSO was used to pick the features to be used in the project. Comparing feature selection with dimensionality reduction, feature selection differs from the latter. However, dimensionality reduction reduces the number of attributes in the dataset by creating new combinations of attributes, whereas feature selection reduces the number of attributes in the dataset by including and excluding attributes already present in the data without altering them. Improvements in Cat Swarm Optimization were made by using two techniques (CSO). A crossover operation was employed in the initial stage of the searching mode, rather than arbitrarily generating N copies of the existing cat (all of which are candidate solutions). The original method was replaced in the second stage of the seeking mode to adjust the position of the cats. As an example, using TF-IDF with ICSO is more accurate than TF-IDF alone when it comes to text classification of large amounts of large-scale text.

To compare the classic KNN method with the enhanced KNN algorithm, Yilin Bel et al. devised an improved KNN algorithm. Weights are assigned to each of the classes as a result of performing the classification in the query instance neighborhood of the standard KNN classifier. Algorithm takes into account class distribution around a query instance to guarantee that the allocated weight does not negatively impact outliers. Using the updated KNN technique based on cluster denoising and density cropping, the traditional KNN is no longer limited in its ability to handle big data sets. On behalf of the classification of health large data, Yihong Li et al. suggested an RBF Neural Network classification algorithm based on manifold analysis and closest neighbor propagation (AP) technique [27]. In order to train the classifier, the standard supervised learning method must do preprocessing and feature selection. It is possible to study the data from a variety of perspectives, such as the standpoint of observation space. Clustering the data set on the basis of neighborhood and the threshold

of the gap between classes, and then deleting the clusters with fewer samples in the class, can help to reduce the influence of some isolated points on AP algorithms. 3231 cases of coronary heart disease, 9628 diabetic cases and 5628 bronchial tuberculosis cases were taken from the city health and family planning bureau as well as several municipal hospitals to create this map. More than 85% of these three data sets can be classified. From improving patient risk score systems to forecasting the onset of disease, to developing hospital operations, machine learning approaches applied to medical and health big data can provide meaningful insights. A consistent performance metric is required to evaluate the effectiveness of machine learning approaches.

VIII. CONCLUSION

To better detect and treat disease, we may use big data in the health sector to improve health profiles and predictor models for specific individuals. Understanding the biology of disease is one of the biggest challenges facing the health sector and the pharmaceutical industry today. A disease can be defined at numerous scales, from DNA to proteins to metabolites to cells to organs to ecosystems using big data. That's what we need to model by integrating big data in biology. Various machine learning approaches were used to discuss the applications, processing, and handling of huge data in this research. Large-scale data is also utilized to measure machine learning models' performance. Through the use of multiple methodologies, machine learning aids in effective decision-making by predicting diseases and making timely diagnoses that benefit the health of patients. Information can be predicted in advance, and diseases can be averted at an early stage if they are detected early. With the use of various algorithms, machine learning allows for the creation of models that accurately predict variables. With the rapid implementation of EHR, fresh data about patients has become available in abundance, which is a huge boon for enhancing human health knowledge. Large-scale adoption and usage of big data utilizing machine learning will be seen in the future in the healthcare industry and in the health sector organization. Large-scale data analytics will become increasingly widespread, which will increase the importance of issues such as security protection, standardization, privacy protection, and governance, as well as enhancing tools and technology. In the health care industry, big data analytics and applications are at an early stage of development, but significant developments in platforms and tools can speed up the process.

REFERENCES

1. You M, Yang G, Chen Y, et al. (2017) A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inf 97: 120–127.
2. Luo G, (2016) Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Inf Sci Syst 4: 2.
3. Zhang Y and Zheng T, (2017) A big data application of machine learning-based framework to identify type 2 diabetes through electronic health records, In: International Conference on Knowledge Management in Organizations, Springer, 451–458.

4. Gerrero-Curieses A, Munoz-Romero S, Bote-Curiel L, et al. (2019) Deep learning and big data in health sector: A double review for critical beginners. Appl Sci 9: 2331.
5. Hao Y, Hwang MCK, Wang L, et al. (2017) Disease prediction by machine learning over big data from health sector communities. IEEE Access, 5: 8869–8879.
6. Manogaran G, Lopez D, Thota C, et al. (2017) Big data analytics in health sector internet of things, In: Innovative Health sector Systems for the 21st Century, Springer, Cham, 263–284.
7. Lopez D and Manogaran G, (2017) A survey of big data architectures and machine learning algorithms in health sector. Int J Biomed Eng Technol 25: 182.
8. Khamlichi KY, Chaoui NEH and Khennou F, (2018) Improving the use of big data analytics within electronic health records: A case study based open HER. Procedia Compute Sci 127: 60–68.
9. Sharma M, Kaur P and Mittal M, (2018) Big data and machine learning-based secure health sector framework. Procedia Compute Sci 132: 1049–1059.
10. Xu Y, Qiu J, Wu Q, et al. (2016) A survey of machine learning for big data processing. EURASIP J Adv Signal Proc 2016: 67.
11. Sisodia DS and Sisodia D, (2018) Prediction of diabetes using classification algorithms. Procedia Comp Sci 132: 1578–1585.
12. Pereira N, Taslimitehrani V, Pathak J, et al. (2015) Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inf 216: 40.
13. Stewart WF, Sun J, Choi E, et al. (2017) Using recurrent neural network models for early detection of heart failure onset. J Am Med Inf Assoc 24: 361–370.
14. Liu Z, Zhang S, Jin B, et al. (2018) Predicting the risk of heart failure with EHR sequential data modeling. IEEE Access 6: 9256–9261.
15. Calvert J, Hoffman J, Jay M, et al. (2016) Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. JMIR Med Inf 4: e28.
16. Hall MK, Pare JR, Venkatesh AK, et al. (2015) Prediction of in hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 23: 269–278.
17. Neapolitan R, Zexian SE, Roy A, et al. (2018) Using natural language processing and machine learning to identify breast cancer local recurrence. BMC Bioinfor 19: 65–74.
18. Garg S and Gupta P, (2020) Breast cancer prediction using varying parameters of machine learning models. Procedia Comput Sci 171: 593–601.
19. Alkhawaldeh RS, Al-Shami F and Al-Shargabi B, (2019) Enhancing multi-layer perception for breast cancer prediction. Int J Adv Sci Tech 130: 11–20.
20. Seenivasagam V and Suijitha R, (2020) Classification of lung cancer stages with machine learning over big data health sector framework. J Ambient Intell Humanized Comput 2020.
21. Olugbara OO and Adetiba E, (2015) Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. Sci World J 2015: 786013.

22. Peter T, Delzell DAP, Smith M, et al. (2019) Machine learning and feature selection methods for disease classification with application to lung cancer screening image data. Front Oncol, 9: 1393.
23. Nartowt BJ, Hart GR, Deng J, et al. (2019) Stratifying ovarian cancer risk using personal health data. Front Big Data 2: 24.
24. Hassanat ABA, (2018) Furthest-pair-based binary search tree for speeding big data classification using K-nearest neighbors. Big Data 6: 225–235.
25. Hung JC, Lin KC, Zhang KY, et al. (2016) Feature selection based on an improved cat swarm optimization algorithm for big data classification. J Supercomput 72: 3210–3221.
26. Bei Y and Xing W, (2019) Medical health big data classification based on KNN classification algorithm. IEEE Access, 8: 28808–28819.
27. Li Y and Jiang C, (2019) Medical health big data classification based on KNN classification algorithm. IEEE Access, 7: 176782–176789.
28. Shah NH and Callahan A, (2017) Machine learning in health sector, In: Key Advances in Clinical Informatics, Academic Press, 279–291.