International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (Aug-2021)

http://dx.doi.org/10.12785/ijcds/100168

Diagnosis Support System for General Diseases by Implementing a Novel Machine Learning Based Classifier

Vikas Kamra[1], Praveen Kumar[2] and Masoud Mohammadian[3]

[1]Research Scholar, Amity University Uttar Pradesh, Noida, India
[2]CSE Department, Amity University Uttar Pradesh, Noida, India
[3]Faculty of Science and Technology, University of Canberra, Australia

Received 30 August 2020, Revised 28 March 2021, Accepted 3 April 2021, Published 5 Aug. 2021

Abstract: Millions of folks around the earth are affliction from late disease identification and diagnosis. An incredible amount of health information has been obtained by the latest technologies in digital medical services and information communication technologies. Disease diagnosis and artificially intelligent decision support systems have drawn tremendous attention from many scientists and research community worldwide. Various algorithms developed and applied with the aid of machine learning techniques that can substantially lead to the resolution of the system of health care and can help personnel involved in the early diagnosis of diseases. This research paper will propose an artificial intelligent algorithm which helps us to effectively, rapidly and accurately classify the information. The Proposed Disease Diagnosis Support Systems (DDSS) can assist clinicians to monitor the information, facilitate their evaluation by means of a preparatory treatment and decrease evaluation time per patient. The patients may inevitably be notified and recommended dietary suggestions also. This system will allow clinicians to focus on attending patients in accordance with their homeostasis. It decreases the volume of work of doctors and enables them to define patients who need to be examined more urgently or meticulously. Even with the widespread growth of such systems security of digital data and its privacy is still a major challenge yet to solve.

Keywords: Disease Diagnosis Support System, Machine Learning, Classification Techniques, Artificial Intelligence

1. Introduction

Many of the people worldwide suffer from the absence of early disease recognition or gaps in doctor's understanding of disease and its symptoms. Many algorithms are created and implemented to combat many kinds of illnesses effectively. Machine learning provides the opportunity to evaluate many diagnostic metrics that a doctor may overlook while discussing with a patient for a few minutes only. The aim of Machine Learning is to develop numerical, conceptual, computable, empirical and scientific methods to find patterns to derive understanding from the given information. Machine learning techniques can be used in many distinct fields to transform human life on earth. Healthcare is prevalent with wealthy information and challenging issues, making it a fertile ground for various machine learning algorithms [1].

Machine learning procedures are strong diagnostic techniques. Needless to say, they have to cope with missing information, as this will contribute to information mistakes and restrict the possible patterns and characteristics of clinical decision making. The suggested DDSS model utilizes natural language processing and latent extraction methodology to deal with the problem of missing information projection [2]. As a practical method, disease diagnosis support systems are intended to take complete advantage of patient health record information, influence physicians' clinical judgments via data mining and help them in delivering timely and accurate judgments to patients [3]. The disease diagnosis support system makes the use of advanced medical knowledge and overhauls patient information to facilitate the care of individuals. It aims, from original consultations to a diagnosis and patient progress surveillance, to enhance interactions between patients and doctors. The type and robustness of disease diagnosis support systems differ from each other. In some systems, the user explicitly generates a help application. Certain systems may run semi-actively, but data is only available to the user on request. Some systems can be triggered, executing and displaying data automatically, without waiting for the end user's specific demand [4]. As the domain of machine-based training and learning was progressed, a deeper understanding of theoretical framework for multiple algorithmic methods was developed. In reality, a significant distinction among

E-mail: kamra1984@gmail.com, pkumar3@amity.edu, masoud.mohammadian@canberra.edu.au

machine learning and statistics is that the first one deal with theoretical problems such as computability and generalization while the second one deals with practical problems and their results. Machine learning is mainly used for extracting salient structure of the data that is more useful and knowledgeful than the original data itself. This process is called as feature extraction. Another major use of machine learning is inferring underlying organized class structure, which is also called as classification [5].

Expert systems were created by the AI society in the mid-1960s as a subset of distributed artificial intelligence. The fundamental concept behind expert systems is mere that competence which is translated from a person to a machine. This information is saved on the desktop and consumers are asked to recommend them when necessary. The machine can deduce and reach a definitive answer. It provides guidance and describes the logic behind the recommendation as a natural advisor. An expert system is described as one that includes data collected by a natural specialist and that data is in the context of if-then-else statements [6].

The rule could also be used to conduct inferential data operations to achieve a suitable result. This inference is basically a computer program which offers a framework for the argument and the drawing of findings on the data in the knowledge base. Its application areas are psychiatric treatment, biochemical nanotechnology, agriculture planning, DNA histogram interpretation, knowledge base maintenance etc. Machine Learning based Disease Diagnosis Support System can be of many types. It can be a Neural Network based System. It can also be an intelligent system based on Agent Technology. An expert system can also a DDSS with fuzzy logic implemented in it [7].

Modern healthcare systems have enhanced human life expectancy over the previous two decades. The health care system is probably nowhere near optimal in spite of years of significant advancement. The third major cause of death after heart disease and cancer is a preventable medical error. Effective ways of reducing the incidence of preventable medical errors should, therefore, be implemented. Provided the latest drastic advancement produced by worldwide health care efforts, it is noteworthy that even in advanced nations, human mistakes stay the significant cause of casualties. Medication mistakes are by far the most prevalent category in medical procedures, among all participating variables [8].

In this category of mistakes, it includes inaccurate prescription, wrong medication dose and lack in administration groups. Another typical human mistake, which creates essentially wrong medical choices that result in severe implications, is inaccurate treatment. The adoption of disease diagnosis support system has been long proposed as a mechanism in order to decrease the danger of human error as well as the workload of the medical personnel [9].

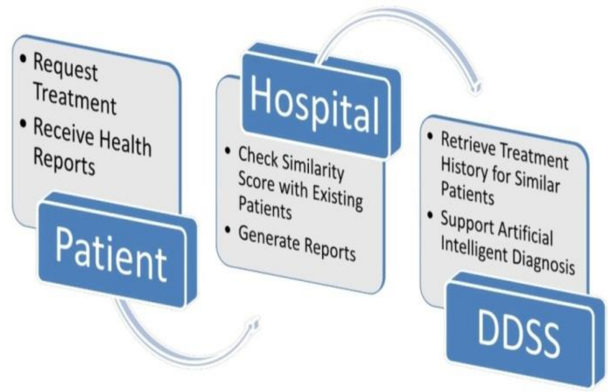

Figure 1. Overview of Communication between Proposed – Disease Diagnosis Support System

The main aim of the proposed disease diagnosis support system is to improve health care procedures by facilitating access to specialist healthcare and decreasing one-to-one visits of patients to the hospitals thus optimizing the time of clinicians in OPDs [10]. The incorporation of remedial knowledge systems and disease diagnosis support systems is one common approach in the application of automated clinical procedures. This method significantly lowers the cost and time needed to enter huge amounts of patient data in each visit and reassembles the diagnostic model dynamically as per the patient's actual variability in time [9].

Machine learning gives a reasonable approach to create advanced, automated, and quantitative biomedical data analysis algorithms. It is the artificial intelligence domain which relies on algorithms that are capable of learning and modifying their framework depending on a series of information observers and also adapt them to an immediate or cost-effective purpose. The advancement of artificial neural networks has precipitated much of the original enthusiasm for applying machine learning into biomedicine [11]. While these neuron speaking claims were largely disproportionate in most cases, one of the important characteristics of ANNs was that, by teaching a certain number of parameters in a connected network of simple non-linear units, they were shown to be capable of simulating any arbitrary role [12].

Clinical diagnosis by its very essence is complicated and sophisticated metacognition and the design of disease diagnosis support system has shown excellent opportunity through soft computing techniques such as neural nets. Substantial savings in lives can be accomplished if people recovering from different kinds of sickness can be diagnosed with accuracy quickly and then adequate therapy can be followed instantly. It is more challenging for credible and strong disease diagnosis support system to help this diagnostic decision process, which is still becoming even more complex, to decrease diagnostic timings and increase diagnostic accuracy [5]. Trying to offer excellent healthcare needs high quality and secure practice. There is continues improvement in the knowledge base of efficient medical treatments, but the medical practice remains behind and mistakes are appallingly common. Disease diagnosis support system provided by data technologies will lastly provide decision-makers with the possibility to attain big efficiencies, narrow gaps between understanding and exercise and also enhanced security, ideally through an electronic medical record as the platform [13].

2. RELATED WORK

Many publications on computer trained methods used in the classification of illnesses were also written, such as Hao Chan associated with Yale University described Melanoma, a skin cancer by the help of deep neural networks. He had an exactness that matched with a dermatologist practicing for many years. Researchers of various reputed universities partnered and demonstrated that treatment of Parkinson's disease can be controlled via Convolution Neural Networks [1]. The case-based reasoning method performs best when applied in complex disease prediction problems because of its automated learning from the instructions available. Researchers [13] provided a scheme for the analysis and sequential measurements of stress-related illnesses by using CBR Technique. Algorithms that provide additional data can merge various scenarios with significant features. However, the shared impact of multi-domain knowledge must be taken into account. These strategies provide theoretical and methodological recommendations for inherent feedback [3]. In order to predict comparable illnesses in patients, Researchers [10] made risk forecasts using a user based CF to discover similarities in clinical characteristics. Researchers [4] provided one-class cooperative filtering strategies based on favorable cases, using the implicit input for binary information to resolve the suggestion issue. Researchers [14] described a random graph for implicit collective filtering to combine Bayesian generative, and developed a stochastic model for estimation. All training data sets need to be loaded into the memory for classical decision tree based classifiers. When the data sets are redistributed, these nodes should rebuild fresh tree structures rather than update old ones.

The Hoeffding Tree algorithm has been produced by Domingos and Hulten, an exponential tree-level methodology to learn from consecutive information streams [9]. Researchers [15] has implemented new profound recurrent network architecture for automatic identification of patients with particular scalp EEG seizures. He suggested a mapping which would concurrently enable the suggested medico clinical characteristics of EEG cranes. Researchers [16] introduced a profound network and the model based on auto encoders to conduct uncontrolled functional education. To characterize the features of public datasets, Convolution Neural Network was used by them. Researchers [17] instituted a paradigm for predicting heart disease by utilizing data mining algorithms that include classification techniques like Decision Tree and Naïve Bayes. The assessment experiments were undertaken to predict the classification's precision. The study shows a different decision tree than the classification of the given dataset by Naïve Bayes. Bayesian network architecture, called internal layer causal modeling, has been created by Researchers [18] along with a Treatment Feature Interactions algorithm that learns the interactions necessary to the model from the information. The probability that metastasis will be safe in five years was comparable by a five-time cross-validation assessment [19]. A latest research tests machines' use of minimally invasive methods to prevent coronary artery illness. Researchers studied supervising machine learners and showed that 88.4% precision relative to multinomial logistic regression, furious unordered induction of rule and C4.5 algorithms was the highest results perceptron in the neural network perceptron [20].

In terms of the variations between facts and practice, Researchers [12] assessed a sequence of released directives and discovered that the implementation of such rules in routine exercise lasted roughly five years on average. Researchers [21] have shown that support vector machines are capable of appropriately classifying genes within functional classifications based on gene data from DNA microarray genetic recombination studies. Researchers [22] introduced an AptaCDSS-E scheme for medical diagnosis decision support in the classification domain for disease forecast with data received from Aptamer Biochip. The scheme used four distinct machine classifiers, combining the forecast outcomes in a group of classifications and produces additional data for the prediction of illnesses. Researchers [23] also pointed out that a crucial consideration is the therapeutic reliability of the disease diagnosis support system: how useful can it be before universal implementation? Some screening is obviously needed since the use of the disease diagnostic support systems in hospital treatment can produce unexpected impacts. It is also indicated that most established schemes are still not prepared for widespread use. Most studies failed to include enough groups or individuals to identify changes in clinical results for sufficient statistical strength. Researchers [24] discovered that the use of the support systems for the diagnosis of diseases greatly reduces the numbers of medication faults. Researcher [12] stated that machine-based learning techniques arose as a crucial field to provide tools and methods for the analysis of biomedical sciences' large volume and multi-directional data. The nervous system of the human brain is the fastest and powerful learning device we understand till date. Because of this, neuroscience has become a subject of the computer teaching society in an effort to recognize fresh concepts and architectures that might account for the outstanding skills of brain-based teaching devices. Researchers [25] concluded that several characteristics have been strongly linked to the capacity of decision support systems to considerably enhance patient treatment. Clinicians and other stakeholders should introduce assistance schemes for clinical choice making that integrate such characteristics, when practicable and suitable. The PCA-LSSVM was implemented by Researchers [26] to diagnose ECG Arrhythmia and the most precise method of teaching was assessed. The ECG Arrhythmia dataset has been experimented to fully automatically diagnose cardiac arrhythmia.

By researching more than 300 deep learning study articles Researchers [29] answered that profound training and deep learning techniques are widely used in the medical sector. Medical diagnosis is performed in deep learning nets with applications, including identification, prediction, segmentation, and classification. It is also discovered that CNN is the most commonly used technique of deep learning and MRI data was used most often in the practice and experimentation of machine learning. Researchers [6] found that chronic neurological dysfunction that affects three percent of the overall population is obsessive compulsive disorder. Researchers [11] said that mental illnesses are a very common group of behavioral conditions that cause a major societal strain. Researchers [2] explained smooth frame rate insights for model learning and choice making generated by improving learning principle. Bipolar disorder is defined by Researchers [8] as a form of seizure disorder referring to an illness with both optimistic and stressful periods. Researchers [30] concluded that the strong combination of mental health conditions in the age cohort reveals the brain as a central organ for the emergence of serious psychological illnesses.

Researcher [7] hypothesized that knowledge - based systems methodologies appear to evolve towards specialist advice, and that a challenge-oriented area is the creation of expert system applications. In different social science approaches, including psychiatry and human behaviour, an expert system may also be used as another form of approach. Researchers [5] proposed scheme of MLP decision-making assistance can attain elevated precision in treatment, above 90%, and comparatively tiny distances under 5% which can be useful to assist the clinical decision-making method for cardiac disease. Researchers [27] have published the latest literature on evolving methods for computer training, optimization protocols and intelligent health apps. Pressing issues have been discussed, including security, pilot research, and actual design and communication between the analytics of information and physicians. Allen Daniel Sunny analyzed that Apriori Algorithm relying on frequent items for machines learning applications provides superior outcomes over Naive Bayes Classifier [28].

Researchers [31] found that there seems to be no real computer program that recognizes or offers an interpretation of the terminology used throughout the psychologist's clinical texts. In addition, Researchers [32] gave new indications of thalamic defects in adolescent OCD and recommended a decrease in the thalamic amount in conjunction with paroxetine therapy. Researchers [33] investigated that software's forecast accuracy between 68 percent to 70 percent which varies from human suicide tempters to non-tempters. The recognition rates of the system were assessed by Researchers [34] for 20 abductions of each object. This verified the efficiency of the framework. Researchers [35] studied that healthcare workers such as children and young adult psychiatrists are often required, in the face of uncertain psychiatric disorders, to make diagnostic and clinical choices in ambiguous conditions with inadequate and trying conditions.

Researchers [36] said that diagnostic choice-making mechanisms are important methods for diagnosing complicated diseases by practitioners. Researchers [37] developed a machine-learned model for the prognosis of main neurological disorders. For detailed Electronic Health Record trials as well as for the patient diagnosis of Autism Spectrum Disorder. Researchers [38] suggested automatized systems for the sample collection. Researchers [39] found that across various fields of practice, disabled people having mental health problems have advised staying away due to societal issues.

3. Modern Healthcare Technologies

Modern healthcare technologies have dramatically improved the human lifespan by advancing medicines, health services and the retention of patient records. Due to advancements in Electronic Health Records (EHR) disease diagnosis support systems attracted significant research interests. The progress of techniques like medical sensors, Cloud and fog computing and IoT has brought unexpected attention to clinical decision support and disease diagnosis systems [13]. Scientists have made considerable attempts to propose fresh medicare strategies, algorithms, technologies, and operating systems. Health care system refers to improving wellness by preventing, treating and examining physical harm, emotional harm, accident, and disease. As computation capacity and wellness information access rises rapidly, artificial intelligence can be used to provide healthcare apps and thus become intelligent [27].

A big amount of disparate instruments are used to control and record the physiological inputs such as glucose level, hypertension, body temperature, blood pressure etc. of distant consumers. Only in monitoring the patient's health status useful perspectives might be acquired from transmissions based on a clinical device [30]. In the recent past, medico-clinical devices have been designed for notifying patients or their household staff during an urgent situation. A huge quantity of medical data is also generated and obtained by patient's healthcare suppliers which help in decision making process of disease diagnosis support systems [13].

Nursing homes use electronic health record mechanisms for the compilation of vast amounts of data on a daily basis. This information from preceding patients plays an essential part in the linear extrapolation treatment results and prediction. This knowledge is used to forecast possible outcomes for each and every person in various mathematical patterns depending on empirical technology. There are Disease Diagnosis Support Systems for several illnesses which analyze previous information from comparable patients and offers personalized therapy suggestions [15]. The support technologies for disease diagnosis have demonstrated improved methods of prescription, decrease major mistakes in medicines, enhance the provision of preventive services and enhance compliance with suggested standards in treatment. These schemes have been shown to be more efficient and to lead in sustained changes in clinical exercise compared with other methods for improving. However, aid technologies for clinical decision-making do not always enhance clinical activity. In the latest research, 72% of clinical exercise has enhanced considerably, but 28% has not [25].

Cloud Computing is used for storing and processing of clinical data to be processed in an artificial intelligent disease diagnosis support system. This intelligent system stores patient's all information and uses it for future reference when the patient visits the hospital again. Nowadays cost and time of hospitalizing the patients are increasing, so IoT based healthcare devices are gaining more attention [31]. In order to ensure the accuracy of the disease prediction and to reduce diagnostic time and cost, this intelligent system has received greater attention. All medical data is stored and processed on the server. But it is a difficult job to safely manage patient health data from unauthorized access. In the past decade, disease diagnosis support systems and health management systems by using digital records have gained very much popularity [14].

Wearable medical devices can gather health-related signals all the time when patient is using them. We can integrate the data received from wearable medical devices with disease diagnosis support systems. The prediction of diseases can be done extensively with the help of machine based learning algorithms as well as data warehousing and mining techniques [13]. Two groups of medical sensors can be classified as IoT-based sensors: a) Sensors based on Physiological Activities, which are used to evaluate basic indications, such as Electrocardiogram, body temperature or heartbeat. b) Sensors based on Bio Kinetics, which collect various body symptoms like velocity fluctuate or movement in or out of the human body. Different sensor types can be reinforced with body parts to provide helpful environmental, temperature, light, and humidity information.

The supervision of the patient status is carried out by all these instruments. These sensors provide important information for the treatment, therapy and follow-up of a patient's behaviors [32]. The remote surveillance scheme is regarded as successful if the big information generated by the medical records or the clinical trials of the patient are effectively evaluated. This diagnostic background is used to assess which measures are required to be taken to prevent the disease. Most remote control systems do not have the efficiency needed to effectively cope with Big Data [19].

A. Machine Learning and Artificial Neural Networks

Machine learning is an interactive intelligence software technology which allows users to discover and enhance their knowledge inevitably while not being specifically programmed. The main focus of this technology is on developing software programs that navigate and use the information for them. The machine learns by observing information and learning by its own experience [33]. This learning process helps machine to take better decisions in the future by using its own experience. The main idea is to permit machines to automatically discover and modify activities without human interference or help [34].

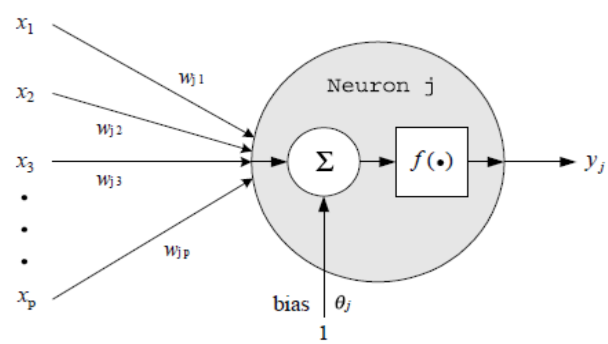

Figure 2. The fundamental structure chunk of ANN, Neuron

Machine Learning allows for huge data analyses. Supervised Learning is the first type of ML Technique. Semi-supervised Learning is the second type of ML Technique. Reinforcement Learning is the third type of ML Technique. Unsupervised Learning is the fourth type of ML Technique [35]. Although the findings are usually quicker and more precise to define lucrative possibilities or harmful hazards, extra effort and resources may also be required to correctly train them. Machine learning can produce it even better to handle big quantities of data by integrating it with artificial intelligence and cognitive computing [5].

An artificial neural network comprises many artificial neurons connected to a particular structure of the system. Inputs are transformed into significant outputs by the neural network. e.g. processing sensory data, like the motions of robots or recognition of a particular item, like a recognizable image [36]. The fundamental structure chunk of an Artificial Neural Network, which is knows as Neuron is revealed above in figure 2. Here, xi is the input values to the neuron and wji is the connection weight, which is in between the different input values. The output of the given Neuron j is represented by yj. Θi is the bias to the Neuron. The activation function f(.) is used for the given Neuron [5].

B. Classification Techniques in Machine Learning

Classification is a supervised learning technique used in machine learning algorithms, in which the computer algorithm derives from the information provided and utilizes it to identify fresh observations. Some of the algorithms for classification are: Logistic Regression, K-Nearest Neighbor and Decision Tree [37].

Logistic Regression is a classified computer training algorithm. In this technique a logistic function models probabilities that describe the feasible results of a given test. Logistic regression is only intended for classification and is particularly helpful to know the impact of multiple independent variables on the individual output vector.

The logistic regression function on P(Y=1|X) is given as:

$$P(Y_n = 1 \mid X_n) = f(AX_n + b) \tag{1}$$

Where:

$$f(X) = \frac{1}{1 + \exp[-(AX + b)]} \tag{2}$$

The error in logistic regression function can be defined as:

$$Error = -\sum_{n=1}^{N} Y_n \log f(X_n) + (1 - Y_n)\log(1 - f(X_n)) \tag{3}$$

The (K) - Nearest (N) Neighbours (N), a kind of passive learning technique because it does not aim to develop an overall inner model but merely maintains training information data. Classification shall be calculated by a simple majority vote of K-Nearest Neighbours of each node. The K Nearest Neighbour Technique is easy to enforce and efficient if training data is big. The Distance Weighted KNN function can be defined as:

$$\hat{f}(x_q) \leftarrow \underset{v \in V}{\arg\max} \sum_{i=1}^{k} w_i \delta(v, f(x_i)) \tag{4}$$

Where:

$$w_i = \begin{cases} \dfrac{1}{d(x_q, x_i)^2} & if \quad x_q \neq x_i \\ 1 & else \end{cases} \tag{5}$$

Decision tree generally creates a tree-like structure for various classification or logistic regression models. It generates a series of rules which can be used for the classification of information when a set of variables and their types are given. A decision tree is easy to comprehend and decipher. Small information processing is required in it and numerical and categorical data are process able by it. The Entropy Function in a Decision Tree Classifier is revealed below:

$$E(n) = \sum_{i=1}^{c} - p_i(c = c_i \mid n) \log_2 p_i(c = c_i \mid n) \tag{6}$$

The Information Gain in Decision Tree Classifier is revealed below:

$$G(n, A) = E(n) - \sum_{v \in Value(A)} \frac{|n_v|}{|n|} E(n_v) \tag{7}$$

4. RESEARCH GAPS IDENTIFIED

Many research gaps have been identified while doing the literature survey in the area of proposed research. Some of them are listed here. The current systems of assistance for disease diagnosis and decision making are always clinically limited. CDSS has extremely restricted access to the medical condition of the patient after discharge. This leads to a number of shortcomings. As a huge number of patients confined to various tests and diagnosis, millions of multifaceted proceedings are increasing quickly in the size of the medical database. The low competence of information systems becomes a new issue for data analyses. The non-uniformity of diagnosis provided by physicians is also a big confront. There is still quite a big standard divergence in doctor's prescriptions. All previous symptoms of the disease, which is a major information repository for doctors to make decisions, cannot always be remembered by the patient [14]. No completely standardized nomenclature and data format have been developed across all EHRs; heterogeneous EHRs can misconstrue health information [9].

Only one disease is predicted at one time with the help of current disease diagnosis systems. We can improve the performance of health care by increasing information accessibility. Healthcare records provide useful simulation and analytical knowledge that can eventually enhance medical treatment and studies. In order for individuals and health organizations to agree to the disclosure and sharing of information, privacy must be shielded infringement. Given the need to predict several illnesses simultaneously, we strive to develop a new model for forecasting many illnesses together. Also, the primary goal is to improve disease prediction with greater precision [13]. A study says that around 60% of physicians understand that the broad application of digital records will contribute to further rob or wasted private information, and about quarter of individuals believe the protection of medical information is not guaranteed. Medical information should preferably be available only if safety and protection are ensured to approved groups or government organizations. In order to conduct information assessment, statistical assessment, and apps to computer pedagogy, scientists require medical information [27].

Data are being gathered and deposited at a large scale in databases. When the quantity rises, the difference between information generation and collection is expanding which is difficult to understand. The huge quantity of coarse data that is accessible is one of the key issues of the internet age. A number of methods recognized as data mining or knowledge discovery are being created to bridge this gap in information [38]. The expansion of information can be characterized as the creation and sharing of knowledge found to individuals in an understandable manner. Medical information to be examined in smaller moment and more detail are provided by classification schemes which are used in medical policy [26]. The pluralities of average systems for diagnosing diseases are usually based on patient symptoms or on information from easy medical surveys. There is at present no disease diagnostic system for extensive diagnoses and feasible biomarker mining with a multi-classifier set [22].

5. Points To Remember While Designing Proposed – DDSS

- It is obligatory that given system is fast in processing. Even if the decision support of the system is magnificent but if it takes too much time to process, it will be worthless.
- The system should foresee the requirements of the physician and provide them with data when it is necessary. The system must satisfy the requirements that were not deliberately realized.
- For a doctor to do the correct decision, the system must be simple in nature. Before the system is installed, usability tests shall be performed [39].
- If you can't fit a single screen result, doctors won't be glad to use it.
- The system regularly updates data that is not already in the system and can be acquired from the physician only, in order to provide efficient decision aid.
- To ensure effective decision support it is essential to maintain the understanding within the system and handling the different systems components.
- Regularly monitor the impact of the system, receive feedback on periodic intervals and answer to the queries on time [40].

6. Proposed - Disease Diagnosis Support System (DDSS)

The proposed model is an upgrade over the Hopfield Network. Each neuron in proposed model is connected to each other neuron. During training process, each neuron remains hidden since after it is inputted to the system. After hidden training process neurons start giving output individually. After that these neurons are combined together to get final output of the proposed system. Networks are trained to evaluate the neurons in accordance with the desired pattern, which is used to calculate the weights. Individual weights of neurons are not changed after their training. Many patterns of same type are used to train the proposed model so that the system adapts the behavior of the learned pattern as per its stability. Proposed – DDSS is an automated guidance and suggestion providing system which offers patient-centered suggestions for the disease forecast. The features of current patients are matched with the target patient to predict the symptoms of the disease. Similarity can be evaluated based on characteristics like symptoms, age, gender and other in the disease diagnosis support scheme. The proposed system predicts the prospects for the occurrence of the specific disease based on a similitude measurement and produces medication suggestions.

The proposed system operates in several junctures to provide patients with more particular suggestions based on their age and sex. It also provides suggestions and recommendations to the patients based on their inputs given to the system. The proposed system, which is used to provide diagnostic support for various general diseases, is an information administration tools designed to enhance the decision-making process in clinical trials. Each patient's properties are combined with EHR knowledge base, and patient-specific suggestions are generated by software systems.

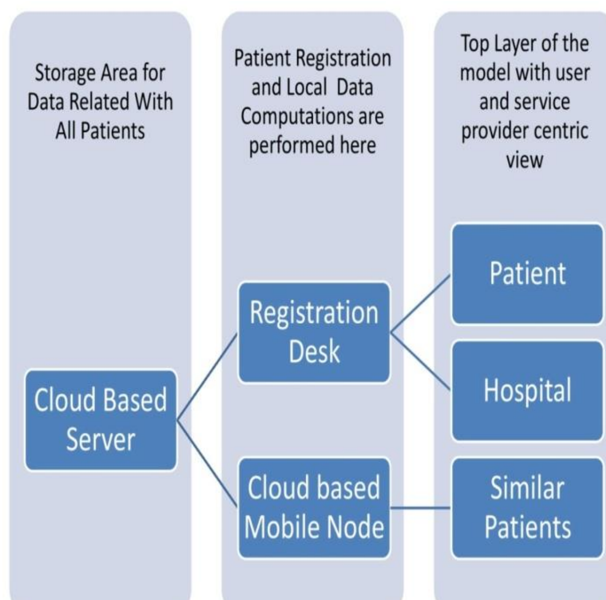

Figure 3. Functionalities Performed at Different Levels of Proposed – Disease Diagnosis Support System (DDSS)

Scikit Learn Module is widely used for this research work, which is a python variant 3 package constructed on the bottom of the sci py database. Scikit Learn is a software database that relies on information patterns, classification and clustering techniques, K-Nearest Neighbor, logistic forest analysis (regression), Classifier based on Naive Bayes Technique, Clustering based on K means Technique, gradient improvement, SVMs and The Decision Tree; including the Python NumPy and SciPy numerical and scientific databases. The following libraries are used for research work: "NumPy" relies on the grid set in n dimensions. "SciPy" is the basic science computation library. The enhanced multimedia console is used by "IPython". "Sympy" offer us the benefit of emblematic math. “Python 3.0” coded the actual implementation.

The list of recommended suggestions is summarized below in Table 1:

TABLE I. LIST OF SUGGESTIONS PROVIDED BY THE PROPOSED – DISEASE DIAGNOSIS SUPPORT SYSTEM

Sr. No.	Suggestion given by the Proposed-DDSS
1	Consume a range of fresh fruits and veggies.
2	Blood pressure should be tested every week.
3	Quit Smoking.
4	Your body fat must be controlled.
5	Meat use should be greatly reduced.
6	It is recommended that you should walk every day for one hour.
7	Core strength training has to be practiced every day.
8	The personal-care and frequent health monitoring guidelines should be maintained.
9	Minimize your diet's use of Salt.
10	Don’t consume coffee empty stomach.

In this computerized system, doctors, medical professionals, and individuals are allowed to directly access the patient's features. Electronic medical records for retrieving clinical characteristics can be requested, too. The physician is sent software-generated suggestions by electronic diagnosis or printouts in the file of a patient. The guidance model is the most significant component of this system.

The objective of this system is to give patients advice on distant aspects. This system is based on suggestions received from specialist physicians on each form of illness and after that it is merged with machine-learning based software.

7. EXPERIMENTAL RESULTS

The proposed model is implemented and executed on Python 3 Platform. Intel Core i3 processor with 8GB RAM is used to perform this experiment on Windows 7.

The parameters used to evaluate the results are Susceptibility, Complexity and Precision values which are compared with other existing classifiers like (K) - Nearest (N) Neighbours (N), The Decision Tree and The Logistic Regression.

The research findings indicate that proposed system can effectively help the registered patient to forecast the likelihood of disease with enhanced forecast precision.

The General Disease Dataset from Kaggle Repository has been taken to predict the disease. Data of 7768 patients was collected in the form secondary dataset and divided in two parts. To train the system 5438 data items were used. The left out 2330 data items tested the system and helped in predicting the results.

Confusion Matrix for the same was generated with TN and TP values as well as FN and FP values. The generated matrix helped out to calculate the susceptibility, complexity and precision values of the proposed model.

The Results of Proposed – DDSS on experimental evaluation are:

a) Susceptibility = 99.12%
b) Complexity = 79.51%
c) Precision = 98.54%

These results are compared with other existing classifiers, which are revealed below in Table 2:

TABLE II. COMPARATIVE TABLE OF PROPOSED – DISEASE DIAGNOSIS SUPPORT SYSTEM COMPARED WITH OTHER EXISTING CLASSIFIERS

Technique / Model	Susceptibility	Complexity	Precision
Logistic Regression	94.34%	76.34%	94.21%
K-Nearest Neighbour	95.7%	79.86%	95.74%
Decision Tree	93.23%	80.30%	93.82%
Proposed-DDSS	99.12%	79.51%	98.54%

The parameters susceptibility, complexity and precision used to empirically evaluate the results of the proposed system have been directly imported from Scikit Learn Module of python variant 3 package. During implementation of the proposed model, these parameters were also applied on Logistic Regression, K-Nearest Neighbor & Decision Tree classifiers and their results have been recorded. No modifications have been done in mathematical equations of these parameters when results of Proposed DDSS have been compared with these classifiers.

The proposed system used certain traits and common characteristics to establish correlation between two patients. These characteristics are allocated by matching the active patient’s attribute values. In comparison to other characteristics, the key characteristics are allocated more weight by the system.

Comparative Graph with names of other existing classifiers on the horizontal axis of the graph and percentage parameter on the vertical axis of the graph is generated to get the Results of Proposed – Disease Diagnosis Support System.

This graph showing the comparison of Proposed – DDSS with other existing classifiers is shown in Figure 4:

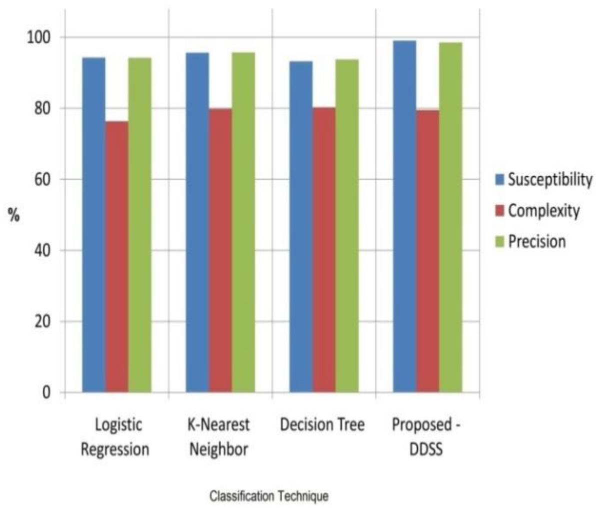

Figure 4. Comparative Graph of Proposed – Disease Diagnosis Support System with other existing Classifiers

Information values received from the source format can be converted into data objects by data transformation and conversion technique. The greatest benefit of the proposed system is its scalability and it will produce more precise outcomes if more people use it. Their results will be added back to the dataset for future evaluation and performance enhancement of the model.

8. CONCLUSION

Early disease forecast and a precise therapy schedule play a key role in leading healthy lives. Since the precision of the healthcare datasets used in this study is high, so a smart disease diagnostic support system is built efficiently for the healthcare industry. The proposed system can help to reduce death loss by an effective early diagnosis of diseases. As different machine learning algorithms have different metrics to input information that affect the efficiency of the system, so it is very important to select effective algorithm to design disease diagnosis support system.

The findings achieved here indicate that the proposed system with 99.12% susceptibility, 79.51% complexity, and 98.54% precision values exceeds over other existing classifiers. The proposed system also boosts medical diagnostic exactitude and lead to superior results for patients. In future too, increased machine computing powers will enable more intricate algorithms to be used with large datasets, which help to design more efficient disease diagnosis and decision support systems for healthcare industry.

ACKNOWLEDGMENT

The authors are indebted to their friends for their cherished support and help. Moreover, the authors still wish to share their gratitude to this paper's editor and reviewers.

REFERENCES

[1] R. Mathur, V. Pathak, and D. Bandil, “Building Machine Learning Based Diseases Diagnosis System Considering Various Features of Datasets,” Emerging Trends in Expert Applications and Security, vol. 841. pp. 357-363, Springer Singapore, 2019.

[2] A. Radulescu and Y. Niv, “State representation in mental illness,” Curr. Opin. Neurobiol., vol. 55, pp. 160–166, 2019.

[3] M. Sun, T. Min, T. Zang, and Y. Wang, “CDL4CDRP: A collaborative deep learning approach for clinical decision and risk prediction,” Processes, vol. 7, no. 5, 2019.

[4] M. Mandzuka, E. Begic, D. Boskovic, Z. Begic, and I. Masic, “Mobile clinical decision support system for acid-base balance diagnosis and treatment recommendation,” Acta Inform. Medica, vol. 25, no. 2, pp. 121–125, 2017.

[5] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, “A multilayer perceptron-based medical decision support system for heart disease diagnosis,” Expert Syst. Appl., vol. 30, no. 2, pp. 272–281, 2006.

[6] D. Denys et al., “Deep brain stimulation of the nucleus accumbens for treatment-refractory obsessive-compulsive disorder,” Arch. Gen. Psychiatry, vol. 67, no. 10, pp. 1061–1068, 2010.

[7] S. H. Liao, “Expert system methodologies and applications-a decade review from 1995 to 2004,” Expert Syst. Appl., vol. 28, no. 1, pp. 93–103, 2005.

[8] Q. Sun, Q. Yue, F. Zhu, and K. Shu, “The Identification research of bipolar disorder based on CNN,” J. Phys. Conf. Ser., vol. 1168, no. 3, 2019.

[9] S. Yang, R. Wei, J. Guo, and L. Xu, “Semantic Inference on Clinical Documents: Combining Machine Learning Algorithms with an Inference Engine for Effective Clinical Diagnosis and Treatment,” IEEE Access, vol. 5, pp. 3529–3546, 2017.

[10] E. Caballero-Ruiz, G. García-Sáez, M. Rigla, M. Villaplana, B. Pons, and M. E. Hernando, “A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs,” Int. J. Med. Inform., vol. 102, pp. 35–49, 2017.

[11] Y. K. Kim and K. S. Na, “Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective,” Prog. Neuro-Psychopharmacology Biol. Psychiatry, vol. 80, pp. 71–80, 2018.

[12] P. Sajda, “Machine learning for detection and diagnosis of disease,” Annu. Rev. Biomed. Eng., vol. 8, pp. 537–565, 2006.

[13] D. Malathi, R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, and A. K. Sangaiah, “Hybrid Reasoning-based Privacy-Aware Disease Prediction Support System,” Comput. Electr. Eng., vol. 73, pp. 114–127, 2019.

[14] H. Yin and N. K. Jha, “A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles,” IEEE Trans. Multi-Scale Comput. Syst., vol. 3, no. 4, pp. 228–241, 2017.

[15] A. Umer et al., “A decision support system for diagnostics and treatment planning in traumatic brain injury,” IEEE J. Biomed. Heal. Informatics, vol. 23, no. 3, pp. 1261–1268, 2019.

[16] M. Sun, F. Wang, T. Min, T. Zang, and Y. Wang, “Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm,” IEEE Access, vol. 6, pp. 77596–77605, 2018.

[17] S. Maji and S. Arora, Information and Communication Technology for Competitive Strategies, vol. 40. Springer Singapore, 2017.

[18] X. Jiang, A. Wells, A. Brufsky, and R. Neapolitan, “A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis,” PLoS One, vol. 14, no. 3, pp. 1–18, 2019.

[19] F. Jabeen et al., “An IoT based efficient hybrid recommender system for cardiovascular disease,” Peer-to-Peer Netw. Appl., vol. 12, no. 5, pp. 1263–1276, 2019.

[20] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman, “Review of Medical Decision Support and Machine-Learning Methods,” Vet. Pathol., vol. 56, no. 4, pp. 512–525, 2019.

[21] M. P. S. Brown et al., “Knowledge-based analysis of microarray gene expression data by using support vector machines,” Proc. Natl. Acad. Sci. U. S. A., vol. 97, no. 1, pp. 262–267, 2000.

[22] J. H. Eom, S. C. Kim, and B. T. Zhang, “AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction,” Expert Syst. Appl., vol. 34, no. 4, pp. 2465–2479, 2008.

[23] A. X. Garg et al., “Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review,” J. Am. Med. Assoc., vol. 293, no. 10, pp. 1223–1238, 2005.

[24] R. Kaushal, K. G. Shojania, and D. W. Bates, “Effects of computerized physician order entry and clinical decision support systems on medication safety: A systematic review,” Arch. Intern. Med., vol. 163, no. 12, pp. 1409–1416, 2003.

[25] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, “Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success,” Br. Med. J., vol. 330, no. 7494, pp. 765–768, 2005.

[26] K. Polat and S. Güneş, “Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine,” Appl. Math. Comput., vol. 186, no. 1, pp. 898–906, 2007.

[27] K. T. Chui, W. Alhalabi, S. S. H. Pang, P. O. de Pablos, R. W. Liu, and M. Zhao, “Disease diagnosis in smart healthcare: Innovation, technologies and applications,” Sustain., vol. 9, no. 12, pp. 1–23, 2017.

[28] A. D. Sunny, S. Kulshreshtha, S. Singh, M. Ba, and H. Sarojadevi, “Disease Diagnosis System By Exploring Machine Learning Algorithms,” Int. J. Innov. Eng. Technol., vol. 10, no. 2, pp. 14–21, 2018.

[29] M. Bakator and D. Radosav, “Deep learning and medical diagnosis: A review of literature,” Multimodal Technol. Interact., vol. 2, no. 3, 2018.

[30] T. Moberget et al., “Cerebellar Gray Matter Volume Is Associated With Cognitive Function and Psychopathology in Adolescence,” Biol. Psychiatry, vol. 86, no. 1, pp. 65–75, 2019.

[31] A. Kamath et al., “Intelligent AI Assisted Psychological Disorder Analysis Using Sentiment Inference,” 2018 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2018, pp. 24–27, 2018.

[32] A. R. Gilbert et al., “Decrease in thalamic volumes of pediatric patients with obsessive- compulsive disorder who are taking paroxetine,” Arch. Gen. Psychiatry, vol. 57, no. 5, pp. 449–456, 2000.

[33] I. C. Passos et al., “Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach,” J. Affect. Disord., vol. 193, pp. 109–116, 2016.

[34] C. Vasios, C. Papageorgiou, G. K. Matsopoulos, K. S. Nikita, and N. Uzunoglu, “Classification of Patients with First-Episode Schizophrenia,” Ger. J. Psychiatry, 2000.

[35] C. A. Galanter and V. L. Patel, “Medical decision making: A selective review for child psychiatrists and psychologists,” J. Child Psychol. Psychiatry Allied Discip., vol. 46, no. 7, pp. 675–689, 2005.

[36] D. Razzouk, J. J. Mari, I. Shirakawa, J. Wainer, and D. Sigulem, “Decision support system for the diagnosis of schizophrenia disorders,” Brazilian J. Med. Biol. Res., vol. 39, no. 1, pp. 119–128, 2006.

[37] M. M. Baig, H. G. Hosseini, and M. Lindén, “Machine learning-based clinical decision support system for early diagnosis from real-time physiological data,” IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, pp. 2943–2946, 2017.

[38] T. Lingren et al., “Electronic health record based algorithm to identify patients with autism spectrum disorder,” PLoS One, vol. 11, no. 7, pp. 1–16, 2016.

[39] D. L. Warkentin, “Psychological Disorders,” Anal. Chem., vol. 65, no. 12, pp. 405–408, 1993.

[40] W. David et al., “Ten commandments for effective clinical decision support : Making the ...,” J. Am. Med. Informatics Assoc., vol. 10, no. 6, pp. 523–530, 2003.

Vikas Kamra received his Bachelor's and Master's Degree in Computer Science and Engineering. Presently he is pursuing his doctoral research in the field of machine learning and data mining from Amity University, India. His research areas include machine learning and data mining.

Praveen Kumar is presently working as Associate Professor in Computer Science & Engineering Department, Amity University, India. His research areas include data mining and e-governance.

Masoud Mohammadian is presently working as Associate Professor in Faculty of Science and Technology, University of Canberra, Australia. His research areas include Computational Intelligence and Intelligent Agents.