*International Journal of Engineering Applied Sciences and Technology, 2018*
**Vol. 3, Issue 9, ISSN No. 2455-2143, Pages 11-16**
Published Online January 2019 in IJEAST (http://www.ijeast.com)

# A DECISION TREE BASED EXPERT SYSTEM FOR MEDICAL DIAGNOSIS

Shweta Taneja, Harsh Goyal, Deepanshu Khandelwal, Abhishek, Aayush Aggarwal
CSE Department
Bhagwan Parshuram Institute of Technology
GGSIPU, New Delhi, India

*Abstract*—**An expert system uses human knowledge to solve complex real world problems. It makes predictions using the given information and the data. An important application of an expert system is in the field of medical diagnosis. In this field, the system inputs various symptoms and predicts the disease. In our proposed expert system, decision tree algorithm is used to predict the disease. The proposed system is implemented and tested using a standard dataset. The results are obtained and accuracy of system is nearly 80 percent. In future, these systems can be very useful for the doctors as well as the patients.**

***Keywords—Expert System, Knowledge Base, Inference Engine, Decision Tree, Indexing.***

## I. INTRODUCTION

Expert systems, commonly known as knowledge-based systems are the most common clinical AI systems. They are designed in a manner such that they are able to analyze the data from a user and provide inference. An expert system is a system which uses the human knowledge captured in a machine and help to resolve problems which otherwise require human intellect. It makes recommendations using the information and the data set provided to it (LIPKIN M, et al., 1958)(Azaab S., et al., 2000). An expert system consists of knowledge base to train itself and learn from the data, Inference engine to use productions by applying some rules and display the result using the User Interface. There are many synonyms of Expert systems like Intelligent Knowledge-Based Systems (IKBS), advice languages, or consultation systems(Beverly G. Hope, et al., 1994).

Medical Diagnosis is a domain which should be performed with proper precision and care. Therefore, in this context machine learning is very useful here as it involves many algorithms of mathematical and statistical analysis (Holman J. G., et al., 2009). In literature, many mathematical and statistical models are used to improve the diagnostic performance of doctors and the treatment of patients.

In our proposed approach, the Expert system is based on decision tree method that leads to a highly accurate diagnosis of the disease in patient's/user's body as per the given symptoms. Our Expert System covers a good amount of diseases and symptoms as it contains a well prepared dataset. This system takes a less amount of time and draws out good results which will be beneficial for both the doctor and the patient.

The overall organization of the paper is as follows: Section 2 gives the Literature survey done in the field of expert system and medical diagnosis. In Section 3, we have given proposed work and the overall architecture. Section 4 explains the working of our system with the help of flowchart, dataset and its working process on a sample input. Section 5 show the final implementation and results. Section 6 states the final conclusion and future scope.

## II. LITERATURE REVIEW

In literature, different algorithms like such as S.V.M, Naïve Bayes, K-Nearest Neighbor etc. are being used in medical diagnosis.

Naïve Bayes, K-Nearest Neighbor and Decision Tree have been used as some of the practices in heart disease prediction so that to cure heart diseases more properly (Soni, J., et al., 2011). Computer Aided Design (CAD) systems have also been used so to detect and predict many medical diseases (Yassin, N. I., et al., 2017). SVM and Random Forest technique is used to detect the diabetes by using iris images(Samant, P., et al., 2018).

Extreme learning machine has also been used in literature for different medical problems (Eshtay, M., et al., 2018). Breast Cancer is one of the most common diseases in women and to detect it, there are many algorithms proposed.These are SVM, k-Nearest Neighbor, Random Forest and Decision Trees which are used in Breast Cancer detection and diagnosis (Eshtay, M., et al., 2018). Liver disease diagnosis is done with the help of SVM and k-Nearest Neighbor in (Hamid, K., et al., 2017). Cardiac Arrhythmia has been classified using SVM, kNN, Random Forest and Logic Regression (Shimpi, P., et al., 2017) and Cardiac Abnormality is detected using SVM in the given research paper (Bhattacharya, A., et al., 2017). SVM, kNN and Decision Tree have also been used to detect Chronic problems in human beings and diagnose it (Anakal, S. et al., 2017). Headaches and Mood Disorders have been cured using Fuzzy Logic (Farrugia, A., et al., 2013) and Bayesian Model (Kim, Y. K., et al., 2018) respectively.

*International Journal of Engineering Applied Sciences and Technology, 2018*
**Vol. 3, Issue 9, ISSN No. 2455-2143, Pages 11-16**
Published Online January 2019 in IJEAST (http://www.ijeast.com)

Diabetes Mellitus is one of the most harmful diseases occurring very commonly in human beings and is being operated using SVM (Kavakiotis, I., et al., 2017)and Naïve Bayes (Maniruzzaman, M., et al., 2016). SVM along with other classifiers is also used in medical diagnosis and biomedical engineering researches (Foster, K. R., et al., 2014).

Adaptive Neural Networks is also one of the machine learning algorithms which is used in the improvement of medical field by providing diagnosis methods (Kononenko, I. et al., 2001). AdaBoost algorithm along with SVM, Random Forest and Naïve Bayes are also used for the diagnosis of Hepatitis and Diabetes (Li, M., et al., 2007). Fuzzy Logic algorithm have also been used to develop self-learning diagnosis systems (Lu, X., 2010).

From the above literature survey, we conclude that Decision tree and Random forest are the best algorithm which can work in medical diagnosis with higher accuracy and speed. So, we selected decision tree as an efficient algorithm which can meet our needs in this work.

## III. PROPOSED WORK

In our proposed work, an expert system consists of a Knowledgebase which consists of the required data, an Inference Engine which applies the production rules to obtain the most accurate results and a User Interface which will help the user to interact with the system. This system will take inputs from the patients or the doctors in terms of the symptom(s) and will predict the most appropriate medical disease according to the data available in the Knowledge Base of the system. The main feature of this system is that it makes use of a Decision Tree Algorithm along with Inference Engine through which it will reach up to a particular disease according to the symptoms given by the user.

### *A. Architecture of Proposed Expert System*

The architecture of proposed system is given in figure 1. It comprises of four major parts namely, Knowledge Base, Inference Engine,Cache(Indexing) and User Interface. Firstly, the system collects the data from the data sets, then it processes the data under Inference Engine and indexes the data using cache so as to set the priority and then provides appropriate results to the user.

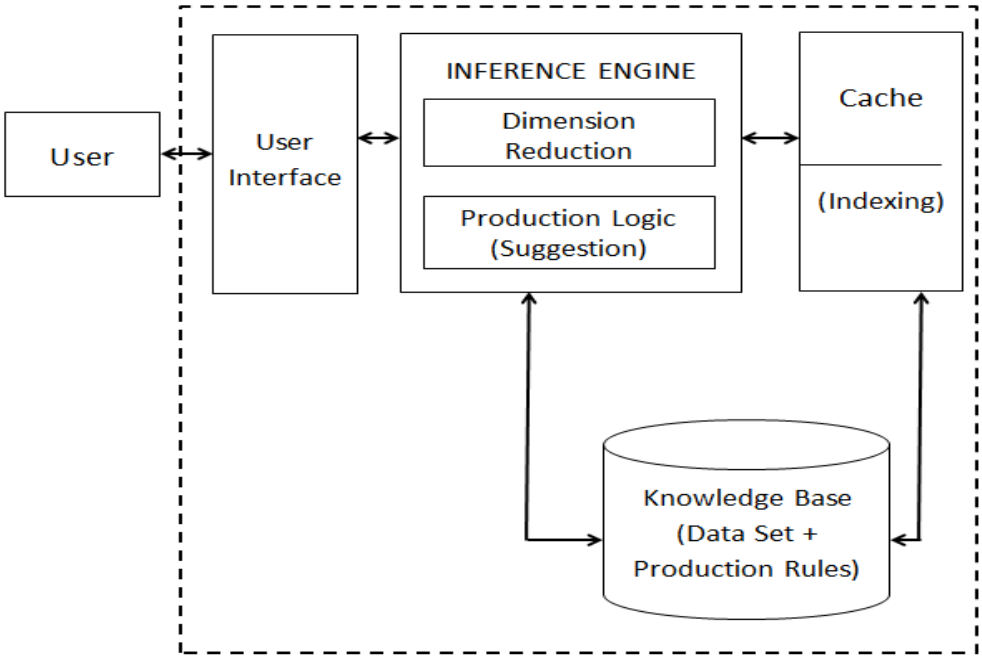


Fig. 1. *Architecture of Proposed Expert System*

*1) Knowledge Base:*The knowledge base of our system contains both heuristic and factual data. Knowledge Base is a place from which the whole system helps itself to train and generate accurate results. The Knowledge Base consists of the data and various production rules which are applied that tell us about specific actions under different conditions or factors.

*2) Inference Engine:* The Inference Engine is the main working component of an expert system. This consists of Production Logic and Dimension Reduction part. The Production Logic helps us to apply highly efficient production rules in the correct order so that no conflicts arise. Dimension Reduction reduces the overall parameters of the dataset by taking queries from the user i.e. the symptoms to reduce the complexity and fasten the process. Firstly, we have full sized dataset on which after taking the inputs from the user, all the diseases which are associated with those input symptoms are selected with all their related symptoms and a new virtual dataset is created

*International Journal of Engineering Applied Sciences and Technology, 2018*
**Vol. 3, Issue 9, ISSN No. 2455-2143, Pages 11-16**
Published Online January 2019 in IJEAST (http://www.ijeast.com)

in the output and then this dataset will be used to fetch results using Decision Tree algorithm. So, the overall dimensions and entries are reduced in the new dataset created at the output.

*3) Cache(Indexing):* The main work of the Cache part is to provide the indexing after the data is fetched. When the data is fetched and cleaned, then the classes are created and to classify them, we provide the priorities or index numbers to each class such that to identify which class should be considered first.

*4) User Interface:* User Interface is an integral part of the system. It helps the user to interact with the system and eases the process of the medical diagnosis from the user's side.

## IV. WORKING OF SYSTEM

Firstly, this system gathers or takes the input from the user in the form of one or more symptoms. After the user inputs the symptom(s), the Expert System will fetch all the diseases which are related to those symptom(s) in our data set or database i.e. those diseases whose symptoms are not present in the user's entered symptom(s) list will be removed for that particular case. After the data fetching phase, the data cleansing takes place which results in the removal of non-usable data. Then the Cache provides indexing and priorities to the diseases whose symptom(s) occur the most which have been selected for a particular case. After all this process, when all the diseases are prioritized and ready for classification, finally the Decision Tree Algorithm is applied which starts from the root or starting index and goes on to ending index and results in the fetching of the final disease to be predicted by the Expert System. So the User Interface will provide the results about that particular case to the user which is in form of diseases. This system makes it easier for users to work upon their health and also helps doctors to analyze well about every disease and its cause of origin. This system in future will be able to give proper reports to the doctors as well as patients about the particular disease and its origin, symptoms and how to cure that disease by proper intake of medicines preferred by the system.

### *A. Flowchart*

The flowchart showing the working of the proposed system is given in figure 2.

The system workflow starts when user interacts with system through GUI. The user gives symptoms of a disease as input to the system. The data input is firstly processed and the processed data is sent to Cache. In Cache, the parameters are checked if they exist previously or not. If the parameters are found in Cache, it will go to Reduce Dataset then it will form Decision tree. If all the parameters were not found in Cache then it will move to next stage known as Actual Dataset.Actual Dataset is the Data set of this System which is already refined by ETL process. The parameters which user enters are thus matched with all possible symptoms present in dataset and all possible diseases from dataset are obtained.

Next step is to feed the reduced dataset into Decision Tree Algorithm which will help us to identify most likely the diseases which can be result of these symptoms. Decision Tree also help to trace all the possible diseases which can be result of these symptoms.

Finally, result is concluded with the help of Decision Tree and the result is showed to user with the help of GUI and also being transferred into cache for the symptoms for faster access of the results.

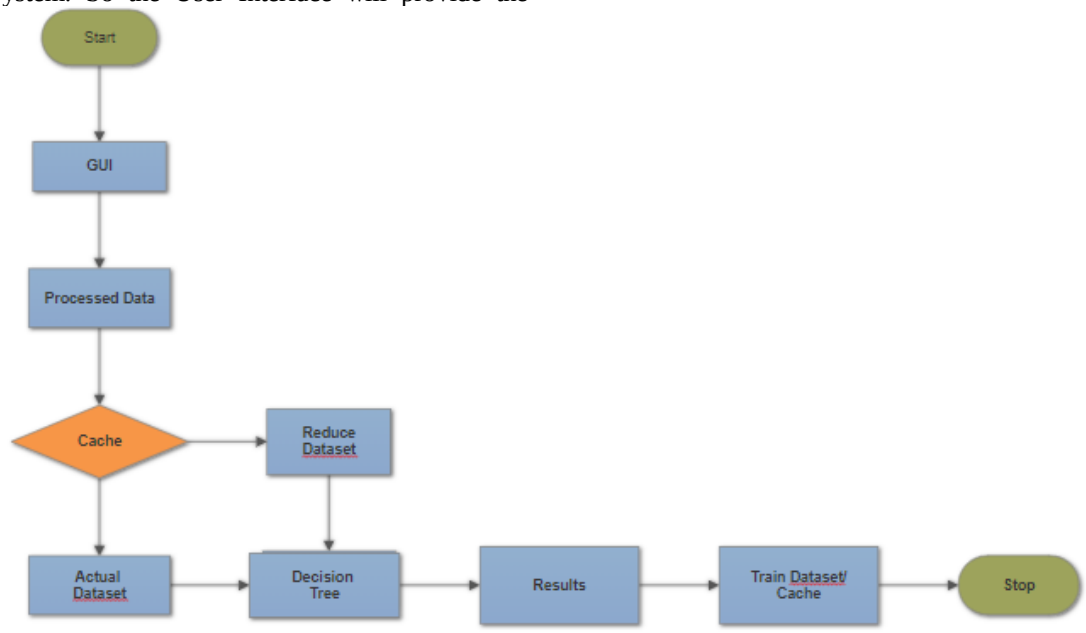


Fig. 2. *Flowchart of Proposed Expert Sys*

*International Journal of Engineering Applied Sciences and Technology, 2018*
**Vol. 3, Issue 9, ISSN No. 2455-2143, Pages 11-16**
Published Online January 2019 in IJEAST (http://www.ijeast.com)

### *B. Dataset*

The Dataset we have taken from Kaggle for the experimentation part (Larmuseau P, 2018). Figure 3 shows the snapshot of the data set.

This Dataset basically consists of two main parameters which are Disease and Symptom. This dataset is a merged dataset and it consists of a disease and its multiple symptoms line by line. There are many redundant disease entries in this dataset but no redundant combination of disease and symptom will be found in the dataset. This dataset consists of a total of 5568 number of rows and 3 columns and a total of 16704 entries.

Originally there were 4 excel sheet in csvform which were linked with the help of disease and symptoms.We combined all sheets to form one master sheet known as database pivoted which is responsible for governing all the result of our system.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Back ache or pain | | | | | | |
| 2 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Flank pain | | | | | | |
| 3 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Abdominal swelling (Stomach swelling) | | | | | | |
| 4 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Low blood pressure | | | | | | |
| 5 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Kidney pain (Flank pain) | | | | | | |
| 6 | 1 | Abdominal aortic aneurysm[(enlarged major blood vessel) | Low back ache or pain | | | | | | |
| 7 | 2 | Abdominal swelling | Swelling | | | | | | |
| 8 | 2 | Abdominal swelling | Retaining fluid | | | | | | |
| 9 | 3 | Abdominal trauma | Trauma | | | | | | |
| 10 | 3 | Abdominal trauma | Lower abdominal pain | | | | | | |
| 11 | 4 | Abrasions[] (scrapes) | Skin trauma | | | | | | |
| 12 | 4 | Abrasions[] (scrapes) | Painful rash | | | | | | |
| 14 | 4 | Abrasions[] (scrapes) | Arm ache or pain | | | | | | |
| 15 | 4 | Abrasions[] (scrapes) | Pain or soreness of breast | | | | | | |
| 16 | 4 | Abrasions[] (scrapes) | Chest pressure | | | | | | |
| 17 | 4 | Abrasions[] (scrapes) | Ear pressure | | | | | | |
| 18 | 4 | Abrasions[] (scrapes) | Hand, finger ache or pain | | | | | | |
| 19 | 4 | Abrasions[] (scrapes) | Headache | | | | | | |
| 20 | 4 | Abrasions[] (scrapes) | Leg ache or pain | | | | | | |
| 21 | 4 | Abrasions[] (scrapes) | Neck ache or pain | | | | | | |
| 22 | 4 | Abrasions[] (scrapes) | Rash | | | | | | |
| 23 | 4 | Abrasions[] (scrapes) | Skin sores | | | | | | |
| 24 | 4 | Abrasions[] (scrapes) | Chest pain | | | | | | |
| 25 | 5 | ACE inhibitor induced cough[blood pressure medication side | Cough | | | | | | |

Fig. 3. *Dataset used*

### *C. Working On a Sample Disease*

For showing the results we selected 4 sample symptoms which are Backache, Headache, Dizziness and Suffocation. These all symptoms are related to Nausea when a person feels very annoying and uncomfortable. Then we obtain the Decision tree shown in figure 4. After traversing one of its branch we reached to final conclusions that is Target Disease. The traversal of decision tree for the given symptoms is shown with help of shaded yellow part which concludes that in leaf node, it is Target Disease belonging to class Medication Reaction. Medication Reaction is a disease which can be caused due to intake of impropermedicines or drugs.Medication Reaction can be very harmful in certain situations as medicines are designed for separate diseases and can cause severe side effects.

## V. RESULT

The Expert System is implemented using Decision Tree and and the results showed the diseases which can be caused by the affection of these symptoms. Here, we took four symptoms Backache, Headache, Dizziness and Suffocation for which we run the Expert System. The system firstly fetched a new dataset containing 136 diseases and their respective number of symptoms with them. Now, this new dataset has reduced the number of diseases from the original dataset and it has also reduced total entries. This reduced dataset only consists of entries or diseases which will be useful for evaluating or applying the Decision Tree algorithm on the input symptoms. Now after fetching out these diseases it will give index and prioritize the diseases and take out all symptoms of selected diseases and recommend or suggest user that he/she has a particular symptom and will finally jump to a disease. This work will be done using production rules and Decision Tree algorithm. So, the doctor or the patient will have clear view of the Decision Tree made and doctor would then be able to traverse to a disease using some help from the patients and his medical tools.

*International Journal of Engineering Applied Sciences and Technology, 2018*
**Vol. 3, Issue 9, ISSN No. 2455-2143, Pages 11-16**
Published Online January 2019 in IJEAST (http://www.ijeast.com)

## VI. CONCLUSION

In this paper, we have proposed an expert system based on decision tree algorithm for medical diagnosis. The proposed system understands, learns and trains the data provided in the data set and apply rules and algorithms to calculate the best possible results. It takes symptoms of diseases as input and predicts the disease. It is implemented using a standard dataset and achieves an accuracy of nearly 80 percent. In future, these systems can be very useful as they can predict the disease faster than the existing systems and can also achieve higher accuracy.

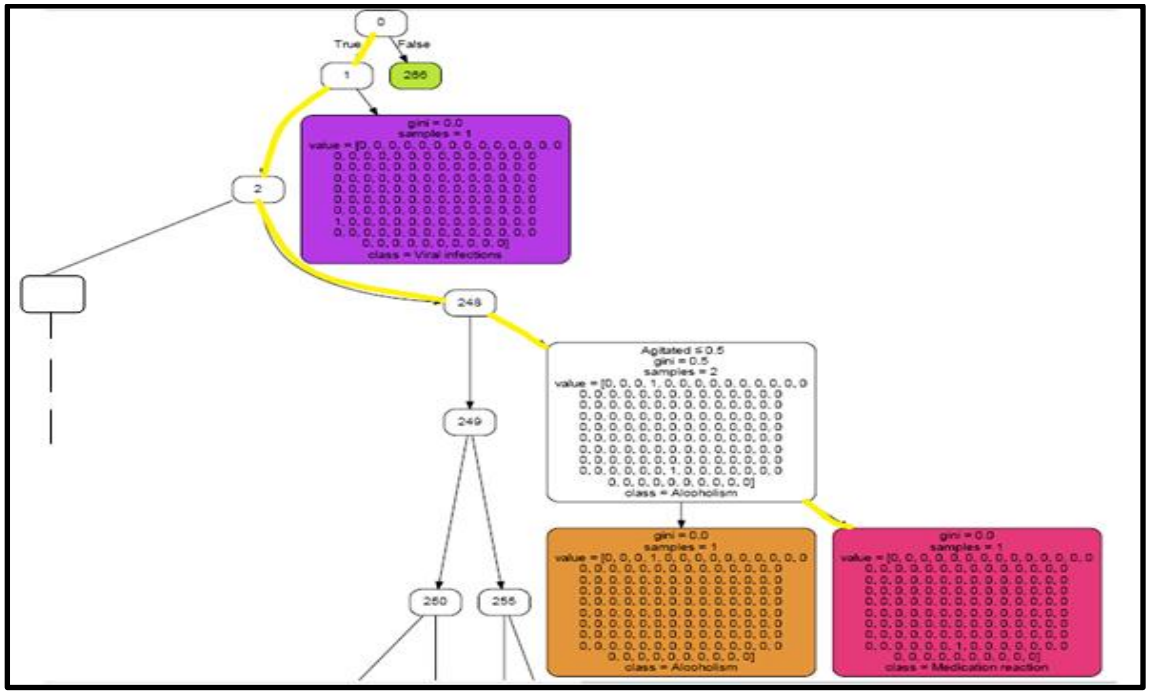


Fig. 4. *Decision Tree formed to traverse Medication Reactio*

## VII. REFERENCES

1. Azaab S., Abu Naser S., and Sulisel O.(2000),"A proposed expert system for selecting exploratory factor analysis procedures".*Journal of the college of education*, 4(2):9-26.
2. Anakal, S., &Sandhya, P. (2017, December). Clinical decision support system for chronic obstructive pulmonary disease using machine learning techniques.In *Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017 International Conference on* (pp. 1-5).IEEE.
3. Bhattacharya, A., Mishra, M., Singh, A., &Dutta, M. K. (2017, November). Machine learning based portable device for detection of cardiac abnormality. In *Emerging Trends in Computing and Communication Technologies (ICETCCT), International Conference on* (pp. 1-4).IEEE.
4. Beverly G. Hope, Rosewary H. Wild (1994), "AnExpert Support System for Service Quality Improvement".*Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Science*.
5. Eshtay, M., Faris, H., & Obeid, N. (2018). Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems. *Expert Systems with Applications*, Sciencedirect, 104, 134-152.
6. Foster, K. R., Koprowski, R., &Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomedical engineering online*, *13*(1), 94.
7. Farrugia, A., Al-Jumeily, D., Al-Jumaily, M., Hussain, A., & Lamb, D. (2013, December). Medical Diagnosis: are Artificial Intelligence systems able to diagnose the underlying causes of specific headaches?. In *Developments in eSystems Engineering (DeSE), 2013 Sixth International Conference on*(pp. 376-382). IEEE.
8. Hamid, K., Asif, A., Abbasi, W., &Sabih, D. (2017, December). Machine Learning with Abstention for Automated Liver Disease Diagnosis.In *Frontiers of Information Technology (FIT), 2017 International Conference on* (pp. 356-361).IEEE.

9. Holman J. G. & Cookson M. J. (2009),"Expert systems for medical applications".*Journal of Medical Engineering & Technology*.
10. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., &Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, *15*, 104-116.
11. Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, *23*(1), 89-109.
12. Kim, Y. K., & Na, K. S. (2018). Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *80*, 71-80.
13. Li, M., & Zhou, Z. H. (2007). Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *37*(6), 1088-1098.
14. Lu, X., & Li, X. (2010, August). The model of medical diagnosis based on machine adaptive learning. In *Natural Computation (ICNC), 2010 Sixth International Conference on*(Vol. 6, pp. 2808-2811). IEEE
15. Lipkin M, and Hrdyj, D. (1958), "Mechanical correlation of data in the differential diagnosis of hematological diseases".*Journal of the American Medical Association*, 166, pp. 113-125.
16. Maniruzzaman, M., Kumar, N., Abedin, M. M., Islam, M. S., Suri, H. S., El-Baz, A. S., &Suri, J. S. (2017). Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Computer methods and programs in biomedicine*, *152*, 23-34.
17. Soni, J., Ansari, U., Sharma, D., &Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, *17*(8), 43-48.
18. Samant, P., &Agarwal, R. (2018). Machine learning techniques for medical diagnosis of diabetes using iris images. *Computer Methods and Programs in Biomedicine*, Sciencedirect, 157, 121-128.
19. Selvathi, D., &AarthyPoornila, A. (2017, July). Performance analysis of various classifiers on deep learning network for breast cancer detection. In *Signal Processing and Communication (ICSPC), 2017 International Conference on*(pp. 359-363). IEEE.
20. Shimpi, P., Shah, S., Shroff, M., &Godbole, A. (2017, July). A machine learning approach for the classification of cardiac arrhythmia. In *Computing Methodologies and Communication (ICCMC), 2017 International Conference on* (pp. 603-607). IEEE.
21. Yassin, N. I., Omran, S., El Houby, E. M., &Allam, H. (2017). Machine Learning Techniques for Breast Cancer Computer Aided Diagnosis Using Different Image Modalities: A Systematic Review. *Computer Methods and Programs in Biomedicine*.Sciencedirect, 156, 25-45.
22. "Symptom Disease Sorting" [Online]. Available: www.kaggle.com/plarmuseau/sdsort/data[Accessed : July 7, 2018