

CHRONIC DISEASES MODELLING – PYTHON ENVIRONMENT**FAISAL MEHMOOD¹, HAROON UR RASHIDKAYANI² AND FATMA HUSSAIN^{1*}**¹*Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan*²*Department of Software Engineering, Lahore Garrison University, Lahore, Pakistan***Corresponding author's email: fatmauaf@yahoo.com***خلاصہ**

ایسے امراض، جو لمبا عرصہ تک چلتے رہتے ہیں، پوری دنیا کی آبادی میں عام ہیں۔ ان امراض سے اموات کی تعداد بھی لاکھوں میں ہے۔ ٹیکنالوجی کے ارتقاء کے ساتھ، مصنوعی ذہانت نے صحت کے محکمے میں ایک انقلاب برپا کر دیا ہے۔ مشین آموزی مصنوعی ذہانت کی ہی ایک شاخ ہے جہاں ایک کمپیوٹر پروگرام کو کسی معاملے میں ایک حتمی رائے دینے کے لیے تیار کیا جاتا ہے۔ مشین آموزی کی مدد سے ایسے ریاضیاتی ماڈل بنائے جاتے ہیں جو کہ انسانی جسم کے اندر پائے جانے والے ماحول کو انسانی جسم کے باہر مصنوعی طور پر پیدا کرنے کی کوشش کرتے ہیں۔ اور اس مصنوعی ماحول پر مریض کی بیماری کی علامات کو مختلف طریقوں سے لاگو کر کے دیکھتے ہیں کہ مریض کو علامات کی بنیاد پر کوئی بیماری ہے اور یہ کہ یہی بیماری انسانی جسم کے اندر رہ کے کیسے پھیلے گی اور کتنا عرصہ لگے گا اس کو کسی خاص سٹیج تک پہنچنے میں۔ ہم نے اس جائزہ میں ایسے ہی مختلف ماڈل کا جائزہ لینے کی کوشش کی ہے جن کو پائیتھن میں بنایا گیا ہے اور جب ان کمپیوٹر ماڈل کو بیمار شخص کی علامات بتائی جاتی ہیں تو یہ ماڈل ان علامات کی بنیاد پر بیماری کی تشخیص کرتے ہیں۔ ان میں سے زیادہ تر ماڈل بیماری کی صحیح تشخیص اور اگلے کچھ سالوں میں اس بیماری کی صورت حال کو کافی حد تک درست پہچاننے میں کامیاب رہے ہیں۔ اس جائزہ میں ہم دمہ کی بیماری، دل کی بیماریوں، الزائمرز، ذیابیطس اور جگر کی بیماریوں کے متعلق مختلف ماڈل کو زیر مطالعہ لائے ہیں۔

Abstract

Chronic infections affect a majority of the population globally. The death toll of chronic disorders is also huge. With the advancement of technology, Artificial Intelligence (AI) can play an important role in healthcare systems. Machine Learning (ML) is a branch of AI, where a computer program is trained to predict the outcome of a decision. Computer simulation models are programs built on the principles of ML and these programs can build computer models that simulate the same physiological environment outside the human body and not only forecast the disease status for upcoming years but also estimate the cost of treatment and the health condition of the patient. We reviewed literature about these computer simulation models built-in Python environment takes in data in the form of attributes of patients and predict the outcome of a disease. Majority of these models have shown significant efficiency and accuracy about predicting the disease with true veracity. We reviewed the prediction models for asthma, cardiovascular disorders, Alzheimer's disease, diabetes mellitus and liver diseases.

Keywords: Chronic diseases, Computer modelling, Machine learning, Python, Predictive models

Introduction

Chronic infections are the ones that have a longer span and usually moderately advancing course. World health organization has categorized chronic diseases into four major types. Those four major types are chronic pulmonary diseases (such as asthma and chronic obstructive lung diseases), cancer, cardiovascular disorders such as stroke and heart attack, and diabetes mellitus (Alwan, 2011). Chronic diseases are the main cause of death and are responsible for 60% mortality rate around the world and 80% of deaths occurring because of chronic disorders were in third world countries. It is estimated that diabetes, chronic respiratory disease, cardiovascular disorders and cancer accounted for 36 million out of 57 million deaths worldwide in 2008 (Alwan et al., 2010). Over the next two decades the number of people who would die due to these chronic diseases, in third world countries is expected to rise substantially (Ezzat et al., 2005).

As chronic diseases progression is slow, it is difficult to conduct clinical trials for such diseases over long periods of time. That is the reason computerized disease models are becoming a better choice for studying chronic diseases. Computers allow the formation of models through various algorithms to study the advancement and progression of diseases. These computer models allow scientists and researchers to predict the patterns involved in the development of disease and forecast the progression of the disease. To achieve this purpose suitable modelling tools are required (Jacob et al., 2010).

Materials and methods:

Systematic review of literature to summarize the past and current role of high-level computer language python in diagnosis, progression tendency and treatment predictions of numerous chronic diseases was conducted. For this purpose, various search engines such as PubMed, Google Scholar, and academic research databases like Web of Science, IEEEExplore and Science Direct were used from September, 2019 to January, 2020.

For computation simulation modelling, the keywords diseases, computer simulation, artificial neural network, python, machine learning, artificial intelligence, chronic diseases, computer modelling, predictive models, asthma, cardiovascular disorders, Alzheimer's disease, diabetes mellitus and liver diseases were used.

Computer prediction models

Doctors are highly specialized in their fields as they have years of experience. Yet they are unable to tell the outcome of the disease. Thanks to the modern advanced computer technology which has made computers to think, although they don't think like a human. Artificial Intelligence (AI) created computer intelligent. Machine Learning (ML) is a branch of Artificial Intelligence (Kononenko, 2001). One or more than one of these types can be used when building an algorithm for a disease prediction model. All these types collect data as the attributes of the patient and find a pattern to draw a conclusion about the diagnosis of disease. These models are given correct and accurate diagnosis data from previous patients and then these models are modified to make a prediction on their own according to the dataset input (Kononenko, 2001).

Designing of computer models have two phases. *Firstly*; a computer model is given some previously tested data whose outcomes are already known to adjust the model so it can make a correct prediction and to check whether it is delivering expected decisions according to data.

Secondly; the model is given the data in the form of attributes of the patients to check the accuracy of the results. Then those results are subjected to verification and validation using different classifiers.

Python

Python is a high-level programming language. It is the most popular and easy-to-use language. It is used widely in healthcare systems. Python provides bulk of open source tools for building ML models that apt for user demands. Python is used in various healthcare fields to provide quick and cheap solutions to the processes and procedures that would otherwise be hectic or costly.

Machine Learning was used in 1970 for the first time to monitor the vaccine dosages given to patients. Until then, we have huge data in healthcare systems on disease symptoms and conditions. Python can be used to build a ML model that uses the previous patient's data and predict the outcome of disease in the coming days or years and can also estimate the cost of treatment.

Python in the diagnostics

According to the National Academy of Sciences, 10% deaths and 6-17% of all complications in patient's disease are caused by the wrong diagnosis of disease. ML can successfully provide a diagnosis based on multiple images techniques such as CT Scan and MRI. The human brain is not efficient at collecting all the perspectives of these images and see them as a whole but ML takes no time to combine, assess and derive a single diagnostic outcome. ML has been very successful in voice recognition, fingerprint recognition, intelligent web searching, driving cars. Since then, scientists have been trying to figure out a way to use them for diagnosis of various diseases (Bart, 2019).

Classification algorithms

Decision tree (DT): Decision tree is a very useful and handy technique for the classification prediction during decision-making process employed by machine learning models. It makes use of a tree-like model for decisions and corresponding outcomes. It is commonly employed to check the outcome of a decision. In DT, decision making involves two steps, the first step is to make root of the tree and second is to make branches of the tree, each branch representing a condition. Decisions or results are represented by leaves of the tree (Martelli and Montanari, 1978).

k-Nearest neighbor(kNN): k-Nearest Neighbor stores all the past data and classify it according to different parameters. Whenever a new case is presented before the kNN, it searches from the past data and compares the situation with previous instances and finds something like current situation (Song *et al.*, 2010).

Naïve bayes (NB): Naïve Bayes is a family of machine learning classifiers, known as probabilistic classifiers. It is not a single algorithm, but a collection of similar algorithms based on the same principle; Bayes' theorem. It states that every single pair of attributes/parameters, that are being classified, is independent of each other. If that is not the case and attributes are not independent of each other then the performance of NB may not be good enough (Lee and Myaeng, 2002; Ling and Zhang, 2002).

Support vector machines (SVM): Support vector machine is a classifier that is based on the principle that two objects are distinct from each other and this logarithm can draw a line between the two objects and classify them as two. That line drawn by this logarithm is known as the threshold and it separates the two different objects. Basically, in the healthcare system, we use this classifier to tell the diseases person and a healthy person apart. We provide patient's data to the algorithm and it makes a decision about the disease (Karthik *et al.*, 2011; Esraa and Mai, 2014).

Model for asthma: Katsuyuki *et al.* (2019) published an article about the development of a model that can predict the onset of chronic pulmonary disease asthma. They assessed the accuracy of their model in predicting asthma using three model classifiers: support vector machine (SVM), logistic analysis, and deep neural network (DNN). Finally, they developed a program model in python using three different classifiers which are DNN, SVM and Logistic Analysis.

Five hundred and sixty-six patients were assessed, 367 patients out of 566 were suffering from asthma and 199 patients were unaffected. Different variables studied during this evaluation were age, sex, cough, wheezing history, repeated symptoms, smoking history, history of allergic diseases etc. Along with these, biochemical finding and lung function test was also considered during calculation. Models that interpreted 10 physical symptoms inputs yielded a 68% accuracy. Models that interpreted on cumulative factors of physical symptoms and biochemical tests generated a 70% veracity and those models which were programmed to include all factors of physical symptoms as well as biochemical tests and lung test returned an accuracy value of 88% (Katsuyuki *et al.*, 2019). All these results are shown in Fig. 2.

Taha *et al.* (2013) studied with kNN, SVM and Random forest test algorithm for the diagnosis of asthma. They obtained an accuracy of 99.34% with SVM, 97.37% with random forest test and 98.70% with K-nearest neighbour.

Model for cardiovascular diseases: According to the World Health Organization, 31% of all deaths worldwide were due to heart diseases and these cause more deaths globally than any other disease. All the hospitals in the world collect data from the patients who encounter heart disease, this data can be used to recognize a pattern of incidence of heart disease. Data mining techniques are useful in recognizing a pattern and to predict the risk factors of heart disease (Animesh and Kavita, 2018).

Shashikant and Chetankumar, (2019) used a model to predict the heart attack in smokers using three machine learning models logistic regression (LR), decision tree (DT) and random forest (RF) test which employed Heart Rate Variability (HRV) for the heart attack prediction. The predictive model was checked using sensitivity, area under curve score as well as accuracy and precision were also taken into consideration. A total of 1562 people were evaluated, of which 751 were non-smokers and 811 people were those addicted to smoking.

Input attributes used for prediction were age, male/female sex ratio, serum cholesterol, maximum blood pressure achieved, resting heart rate, chest pain and fasting blood sugar etc. They tested these attributes using three models and manipulated the data using a python tool. Fig 3. represents a simple scheme of a predictive model

In another study, Ootom *et al.* (2015) designed a model for the prediction of a heart disease known as Coronary Heart Disease (CAD). The disease dataset, obtained from UCI (repository), had 303 disease cases consisting of 76 attributes or parameters. Out of these 76 attributes, 13 were taken into consideration for the prediction model. They used three algorithms bayes net, support vector machine and functional tree for the detection of CAD. WEKA tool was also used for the detection. Accuracy of bayesnet, support vector machine and functional tree were found to be 84.5%, 85.1% and 84.5% respectively. Further studies on cardiovascular disease prediction models are given in Table 1.

Alzheimer's disease prediction model: Das *et al.* (2019) tested a machine learning model which could diagnose the onset and development of Alzheimer's disease. They tested the model named sparse high-order interaction model with rejection option (SHIMR). The model was incorporated with an option that if the model was not confident enough about the disease, it would not give a prediction of diagnosis. In this case, the doctor can diagnose the disease more confidently after invasive and costly pathologies. They obtained data of 151 patients, from Alzheimer's disease Neuroimaging Initiative (ADNI) dataset. They used 14 plasma proteins as the

attributes. These 14 proteins are considered responsible for the advancement of Alzheimer's disease. They also considered the concentration of cerebrospinal fluid (CSF) for the prediction model.

The other classifiers used to compare the results of SHIMR were random forest test (RF), support vector machines (SVM) and decision tree (DT). While comparing these tests, the SHIMR was set at zero-rejection rate. It was observed that with increase in rejection rate, accuracy and reliability of results were improved. They obtained an accuracy (ACC) of 0.74 with zero rejection rate and an ACC score of 0.9 when the rejection rate was increased to 0.38 (Das et al., 2019).

Liver disease prediction model: Liver is the most important organ of the human body and hepatic pathophysiological conditions affect whole body metabolism. There is a great need to detect liver disease at an early stage so that the measures can be taken to treat it properly and avoid complete liver failure which would otherwise result in the death of the patient (Shambel and Pooja, 2018)

JinHoo et al. (2014) worked with various classification algorithms for the detection of liver disease to help physicians in recognition and treatment at an early stage. They employed naïve bayes, decision tree, multi-layer perceptron and kNN.

These programs were compared in predicting the disease according to different parameters like sensitivity, specificity etc., using the Indian liver patient dataset from UCI. They examined 414 liver patients and 165 people suspicious of liver disease. The attributes checked were age, gender, total bilirubin, alkaline phosphatase, alkaline aminotransferase, total proteins and albumin. Classification results in terms of precision were better in case of naïve bayes but decision tree and kNN showed much better classification results in terms of sensitivity.

Tian et al. (2019) inspected a model to forecast the HBsAg (Hepatitis B surface Antigen). They tested the laboratory information of 2235 patients with 4 prediction models which are: decision tree (DT), random forest test (RF), logistic regression (LR) and extreme gradient boosting (XGBoost). Accuracy was checked by the AUC score (area under curve).

They used a total of 30 attributes for the detection model, some of which are: male/female ratio, alcoholic addiction history, HBV previous ancestral history, primary diagnosis, ongoing diagnosis, primary treatment, ongoing treatment, etc. They divided their dataset into two categories of training dataset and testing dataset. Training dataset had a known output, and the model was trained to match the outcome of the prediction. In this study four models, they tested, were built based on python programming version 3.6.

The parameters they used for authenticity of their results were: True positive (TP) which were the positive results correctly recognized as positive, false negative (FN) which were the cases where positive patients were incorrectly reported as negatives, true negative (TN) which were identified correctly as negative, and False positives (FP) which were incorrectly reported as positive. Other parameters were precision, sensitivity, F-score and area under curve (AUC).

Random forest recognized 634 cases as true negative, 31 case as false negative, 4 cases as true positive and 1 case as false positive. Whereas the sensitivity of random forest was 0.95, precision was 0.99 and AUC score with 95% confidence interval was 0.829.

Whereas, according to DT, 627 cases were true negative, 31 cases were false negative, 4 cases were true positive, and 9 cases were false positive. Sensitivity of DT was calculated to be 0.94, precision was 0.97 and the AUC score with 95% confidence interval was 0.19.

Logistic regression model provided with the values of 636, 35, 0 and 0 for true negative, false negative, true positive, and false positive respectively. Whereas it gave the values of 0.95, 1.00, 0.680 for sensitivity, precision and AUC score respectively.

Last model under observation, the extreme gradient boosting gave the best values of all. It provided with 632 true negative, 26 false negative, 9 true positive and 4 false positive cases. It showed a sensitivity of 0.96, precision of 0.98 and AUC score (with 95% confidence interval) of 0.689.

AUC results of XGBoost were best as compared to other models. Results of XGBoost of other parameters were also significantly better than models. So, overall the best model among these four was proven to be the extreme gradient boosting model (Tomita et al., 2019). Some studies conducted on liver disease prediction models are shown in Table. 2.

Diabetes prediction models: Iyer et al. (2015) have experimented with decision tree and naïve Bayes logarithm to detect diabetes. Pima Indian dataset was used for the data used in tests of this experiment. Data mining software that was used was Waikato Environment for Knowledge Analysis (WEKA) Tool. Two methods were employed for prediction: Cross-validation and data percentagesplit method. In later, Data was split into two parts and 70% of data was used for training and 30% data was used for testing purpose. The highest accuracy was shown when algorithms were used with percentage split method.

Sarwar and Sharma, (2012) experimented with naïve bayes logarithm for the diagnosis of diabetes type 2. Data of 415 cases was taken under consideration from various societies of India. About 95% correct diagnosis was achieved with naïve bayes model.

Ephzibah, (2011) constructed a diabetes diagnostic model with genetic algorithm and fuzzy logic. Eight attributes and 769 cases were taken from UCI machine learning dataset. Genetic algorithm selected 3 best attributes out of 8 and then further selection was done by fuzzy logic and 87% accuracy was achieved via this method and a method was deduced with 50% lessercost of treatment.

Table 1: Cardiovascular disease prediction models

Author	Method	Dataset	Results
Elbedwehyet <i>al.</i> (2012)	binary particle swarm optimization, k-nearest neighbor and Support Vector machine	198 samples of heart sound signals, having two classes and 88 attributes	They obtained following results: Sensitivity: 95.24% Specificity: 80% Accuracy: 87.80%
Jabbar <i>et al.</i> (2013)	Genetic algorithm	6 datasetstaken from UCI and one dataset from an Andhra Pradesh Hospital	Accuracy= 95.73%
	k-nearest neighbour		Accuracy= 92.14%
Alizadehsaniet <i>al.</i> (2013)	Naïve bayes	Alizadeh Sani dataset	Accuracy= 94.08%
Yang <i>et al.</i> (2014)	Adaptive-network-based fuzzy inference system (ANFIS) Linear discriminant analysis (LDA)	Korean national health	Accuracy= 80.2%
El-Bialy <i>et al.</i> (2015)	Fast Decision tree	UCI dataset	Accuracy= 78%
	C 4.5 algorithm		Accuracy= 77.5
Acharya <i>et al.</i> (2017)	Convolutional Neural Network (CNN)	ECG Signals	Accuracy=95.22% Sensitivity=95.49% Specificity=94 .19%
Kalaiselvi, (2016)	K-nearest neighbor	UCI dataset	Accuracy: 96.5%
	Naïve Bayes		Accuracy: 94.43%
	Decision tree		Accuracy: 96.1%

Table. 2: Liver disease prediction models

Author	Method	Dataset	Results
Vinceyand Jeba, (2013)	Hidden Markove Model (HMM)	CT Scan Images	This model can detect liver cancer at an early stage.
Dhamodharan, (2014)	Naïve Bayes and FT Tree algorithms.	N/A	Naïve Bayes showed a better performance at predicting Cirrhosis, Cancer and Hepatitis.
Liang and Peng, (2013)	Artificial immune and genetic algorithms	UCI	Higher accuracy and sensitivity value than other classifiers like SVM, Random Forest and Naïve Bayes
Surya and Irshad, (2016)	K-means Clustering with Atkinson Index	Indian Liver Patient Dataset	Better results with Atkinson index than without it.

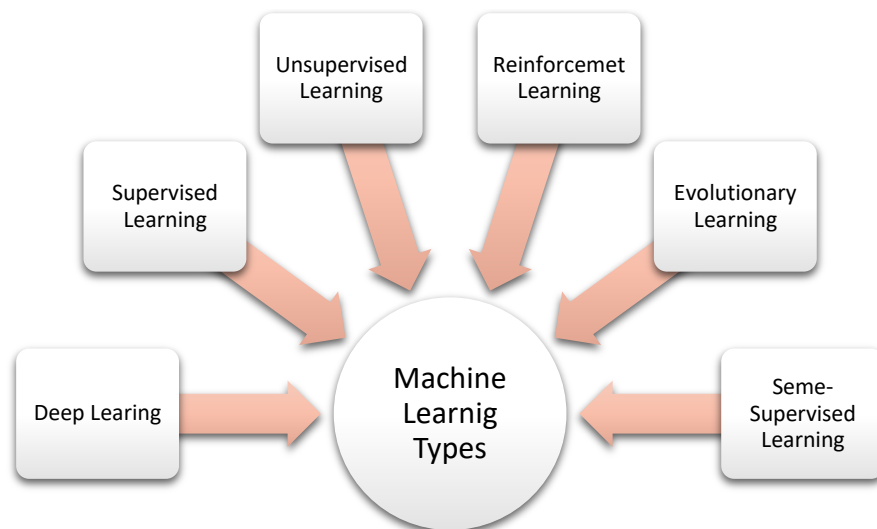


Fig.1. Machine Learning Types

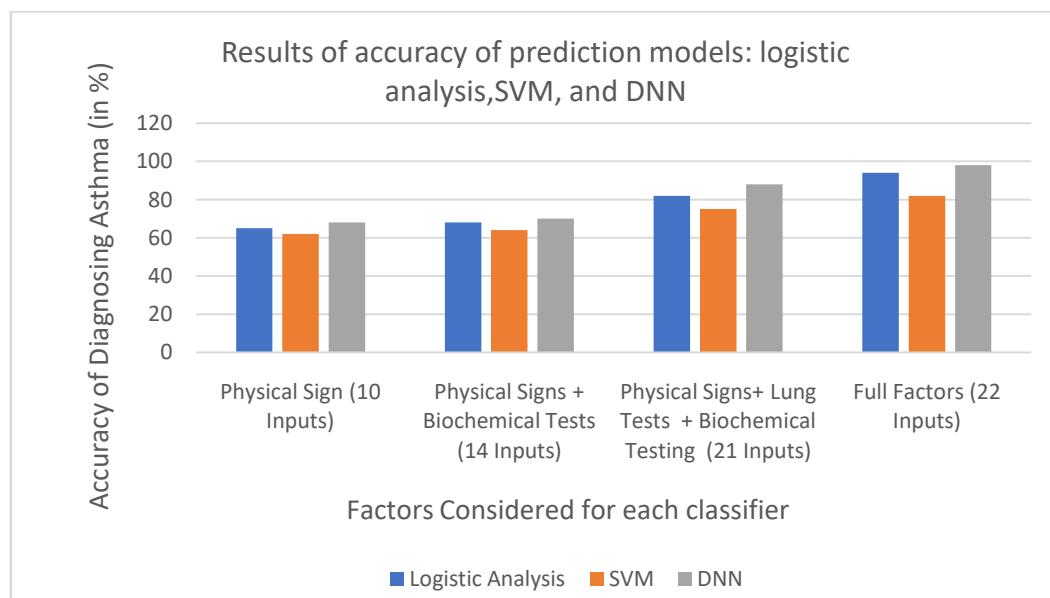


Fig.2. Results of accuracy of prediction models: logistic analysis, SVM, and DNN

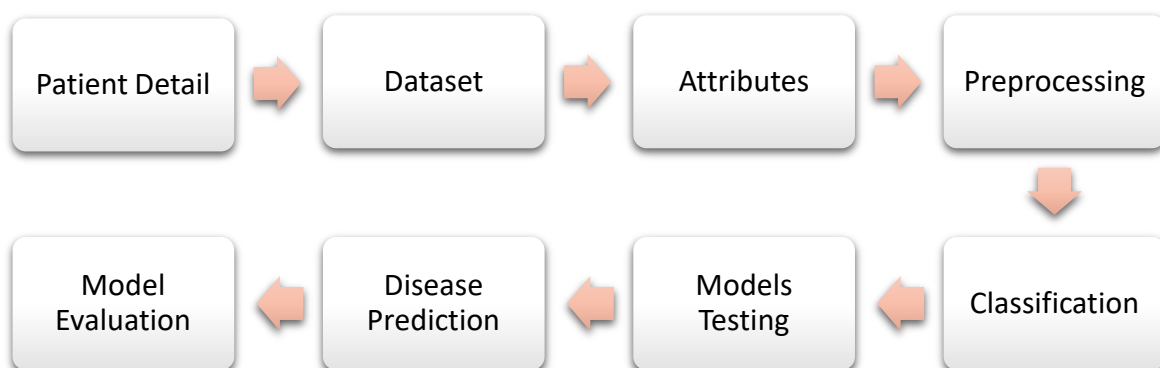


Fig.3. A flowchart of Predictive Model

Conclusion

Past was all about finding ways to cure diseases. Current interest lies in forecasting disease before the disease even develops. We reviewed 5 disease models in this article about diabetes, asthma, CVD, Alzheimer's and hepatitis. They showed significantly accurate results when compared to physiological testing. One important consideration about these models is to train the model and give proper attributes about the patients for disease prediction. According to the results provided by these models, we can say that these models can provide a promising future to the in-silico disease diagnosis in the coming years. There would be no need for the expensive and invasive diagnosis of diseases and just according to the attributes of the subject, a model would be able to predict with better accuracy whether the person has the disease, or the person is healthy. These computer simulations not only provide a timesaving diagnosis but also cut the heavy expanses of invasive treatment. So, with the upcoming technology of computer simulation models, this technology is going to develop more accuracy.

Conflict of interest: The authors declare that there is no conflict of interest regarding the publication of this review article.

References

- Acharya, U.R., H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan and M. Adam. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*. 415:190-198. <https://doi.org/10.1007/s10489-018-1179-1>
- Alizadehsani, R., J. Habibi, M.J. Hosseini, H. Mashayekhi, R. Boghrati, A. Ghandeharioun, B. Bahadorian and Z.A. Sani. (2013). A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*. 111:52-61. <https://doi.org/10.1016/j.cmpb.2013.03.004>
- Alwan, A. 2011. *Global status report on noncommunicable diseases 2010*, World Health Organization. https://www.who.int/nmh/publications/ncd_report2010/en/
- Alwan, A., D.R. Maclean, L.M. Riley, E.T. D'espaignet, C.D. Mathers, G.A. Stevens and D. Bettcher. (2010). Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *Lancet*. 376:1861-1868. [https://doi.org/10.1016/S0140-6736\(10\)61853-3](https://doi.org/10.1016/S0140-6736(10)61853-3)
- Barhak, J., D.J. Isaman, W. Ye and D. Lee. (2010). Chronic disease modeling and simulation software. *Journal of Biomedical Informatics*. 43:791-799. <https://doi.org/10.1016/j.jbi.2010.06.003>
- Das, D., J. Ito, T. Kadowaki and K. Tsuda. (2019). An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ*. 7:e6543. <https://doi.org/10.7717/peerj.6543>
- Dhamodharan, S. Liver disease prediction using bayesian classification. 4th National Conference on advanced computing, applications & Technologies, 2014. 1-3. <https://pdfs.semanticscholar.org/d68d/556dc4f412b0770f8d9c1237685b4e7e5d6c.pdf>
- Dubey, A.K. and K. Choudhary. (2018). A systematic review and analysis of the heart disease prediction methodology. *International Journal of Advanced Computer Research*. 8:240-256. DOI: [10.19101/IJACR.2018.837025](https://doi.org/10.19101/IJACR.2018.837025)
- El-Bialy, R., M.A. Salamay, O.H. Karam and M.E. Khalifa. (2015). Feature analysis of coronary artery heart disease data sets. *Procedia Computer Science*. 65:459-468. <https://doi.org/10.1016/j.procs.2015.09.132>
- Elbedwehy, M.N., H.M. Zawbaa, N. Ghali and A.E. Hassanien. Detection of heart disease using binary particle swarm optimization. 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), (2012). IEEE, 177-182. https://fedcsis.org/archive/proceedings_2012
- Ephzibah, E. (2011). Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis. *arXiv preprint arXiv:1103.0087*. DOI: [10.5121/ijsc.2011.2101](https://doi.org/10.5121/ijsc.2011.2101)
- Ezzati, M., A.D. Lopez, A.A. Rodgers and C.J. Murray (2004). *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, World Health Organization. DOI: [10.1186/1478-7954-1-1](https://doi.org/10.1186/1478-7954-1-1)
- Hashem, E.M. and M.S. Mabrouk. (2014). A study of support vector machine algorithm for liver disease diagnosis. *American Journal of Intelligent Systems*. 4:9-14. <http://article.sapub.org/10.5923.j.ajis.20140401.02.html>
- Iyer, A., S. Jeyalatha and R. Sumbaly. (2015). Diagnosis of diabetes using classification mining techniques. *arXiv preprint arXiv:1502.03774*. DOI: [10.5121/ijdkp.2015.5101](https://doi.org/10.5121/ijdkp.2015.5101)
- Jabbar, M.A., B.L. Deekshatulu and P. Chandra. (2015). Classification of heart disease using k-nearest neighbor and genetic algorithm. *arXiv preprint arXiv:1508.02061*. <https://arxiv.org/abs/1508.02061>
- Jin, H., S. Kim and J. Kim. (2014). Decision factors on effective liver patient data prediction. *International Journal of Bio-science and Bio-Technology*. 6:167-178. DOI: [10.14257/ijbsbt.2014.6.4.16](https://doi.org/10.14257/ijbsbt.2014.6.4.16)

- Kalaiselvi, C. Diagnosing of heart diseases using average k-nearest neighbor algorithm of data mining. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), (2016). *IEEE*, 3099-3103.https://www.aconf.org/conf_71731.html
- Kant, S. and I.A. Ansari. (2016). An improved K means clustering with Atkinson index to classify liver patient dataset. *International Journal of System Assurance Engineering and Management*. 7:222-228.<https://link.springer.com/article/10.1007/s13198-015-0365-3>
- Karthik, S., A. Priyadarishini, J. Anuradha and B. Tripathy. (2011). Classification and rule extraction using rough set for diagnosis of liver disease and its types. *Advances in Applied Scientific Research*. 2:334-345.<https://pdfs.semanticscholar.org/04f4/3017a1bfe17f7ae3f05f683db80910b6d4d0.pdf>
- Kefelegn, S. and P. Kamat. (2018). Prediction and analysis of liver disorder diseases by using data mining technique: survey. *International Journal of Pure and Applied Mathematics*. 118:765-770.<https://pdfs.semanticscholar.org/a0e9/90bc60bc226506984fbb336d836a44c735b7.pdf>
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*. 23:89-109.[https://doi.org/10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X)
- Lee, Y.-B. and S.H. Myaeng. Text genre classification with genre-revealing and subject-revealing features. Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, (2002). 145-150.<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.2500>
- Liang, C. and L. Peng. (2013). An automated diagnosis system of liver disease using artificial immune and genetic algorithms. *Journal of Medical Systems*. 37:9932.<https://link.springer.com/article/10.1007/s10916-013-9932-9>
- Ling, C.X. and H. Zhang. (2002). The representational power of discrete Bayesian networks. *Journal of Machine Learning Research*. 3:709-721.<http://www.jmlr.org/papers/v3/ling02a.html>
- Martelli, A. and U. Montanari. (1978). Optimizing decision trees through heuristically guided search. *Communications of the ACM*. 21:1025-1039.<https://dl.acm.org/doi/abs/10.1145/359657.359664>
- Otoom, A.F., E.E. Abdallah, Y. Kilani, A. Kefaye and M. Ashour. (2015). Effective diagnosis and monitoring of heart disease. *International Journal of Software Engineering and Its Applications*. 9:143-156.https://www.researchgate.net/profile/Emad_Abdallah2/publication/282747944_Effective_diagnosis_and_monitoring_of_heart_disease/links/5bd1a3f7299bf14eac83b058/Effective-diagnosis-and-monitoring-of-heart-disease.pdf
- Sarwar, A. and V. Sharma. (2012). Intelligent Naïve Bayes approach to diagnose diabetes Type-2. *International Journal of Computer Applications and Challenges in Networking, Intelligence and Computing Technologies*. 3:14-16.<https://www.ijcaonline.org/specialissues/icnict/number3/9030-1043>
- Shashikant, R. and P. Chetankumar. (2019). Predictive model of cardiac arrest in smokers using machine learning technique based on Heart Rate Variability parameter. *Applied Computing and Informatics*.<https://doi.org/10.1016/j.aci.2019.06.002>
- Song, B.-H., M.-A. Jung and S.-R. Lee. (2010). A design and implementation red tide prediction monitoring system using case based reasoning. *Journal of Korean Institute of Communications and Information Sciences*. 35:1219-1226.<http://www.koreascience.or.kr/article/JAKO201027935975095.do>
- Tahasamadsoltaniheris, M., Z. Mahmoodvand and M. Zolnoori. (2013). Intelligent diagnosis of Asthma using machine learning algorithms. *International Research Journal of Applied and Basic Sciences*. 5:140-145.https://www.researchgate.net/publication/261870311_Intelligent_Diagnosis_of_Asthma_Using_Machine_Learning_Algorithms
- Tian, X., Y. Chong, Y. Huang, P. Guo, M. Li, W. Zhang, Z. Du, X. Li and Y. Hao. (2019). Using machine learning algorithms to predict hepatitis B surface antigen seroclearance. *Computational and Mathematical Methods in Medicine*. 2019.<https://www.hindawi.com/journals/cmmm/2019/6915850/>
- Tomita, K., R. Nagao, H. Touge, T. Ikeuchi, H. Sano, A. Yamasaki and Y. Tohda. (2019). Deep learning facilitates the diagnosis of adult asthma. *Allergy International*. 68:456-461.<https://doi.org/10.1016/j.alit.2019.04.010>
- Vincey, J. and M. Jeba. (2013). Computer aided diagnosis for liver cancer feature extraction. *International Journal of Engineering Science*. 11:27-30.
- Yang, J.-G., J.-K. Kim, U.-G. Kang and Y.-H. Lee. (2014). Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS-LDA). *Personal and Ubiquitous Computing*. 18:1351-1362.<https://link.springer.com/article/10.1007/s00779-013-0737-0>