# A STUDY ON PREDICTIVE MODELING FOR DISEASE PREDICTION BASED ON SOFT COMPUTING APPROACHES

*Akhil Mathew Philip*[1] *Dr. S Hemalatha*[2]

**ABSTRACT**

Data mining is the process of analyzing data based on different perspectives and summarizing it into useful information. The use of medical datasets has attracted the attention of researchers all over the world. Data mining techniques have been widely employed through a set of medical datasets to develop decision support systems for disease prediction. The knowledge based system can be used as a clinical analytical method to assist medical practitioners in healthcare practice. In this paper a study of different soft computing methods for disease prediction is done.

***Keywords****: Data Mining, Disease Prediction*

## 1. INTRODUCTION

In 1997, the World Health Organization (WHO) found the possibility of the data mining meant for sorting out complications in the medical field. WHO highlights the value of knowledge detection as of repositories of medical data, as it helps medical diagnosis as well as prognosis. A method of determining valuable knowledge from database to shape a form (i.e., standard or outline) that could significantly explain the data is named "Data mining". To determine hidden pattern in data, several machine learning techniques have been used by data mining. There are three main methods used, namely supervised learning, unsupervised learning and semi-supervised learning. Doctors get aided by the Expert systems formed using machine learning (ML) techniques for diagnosing as well as predicting disease. In regression and classification of nonlinear systems, ML is considered as an active realistic method. Systems like this can be immensely multivariate concerning millions of variables. A widespread 'the training dataset' of samples in ML is formed including as much system parameter space as possible [1-2].ML in the field of computer, statistics, AI as well as information theory, is a multidisciplinary technology, which is favored mainly by data scientists intended to use information not directly visible in Big Data.

The distinctive properties of Big Data – enormous, high dimensional, varied, complicated otherwise unstructured, lacking, noisy and incorrect comprise really challenging conventional arithmetic methods, which are mostly suggested for analyzing moderately insignificant samples [13].Over a period, soft computing takes up the important role of computer-aided disease-identification in the opinion of a physician. Some of the main soft computing criteria are error headings, malleability, partial certainty as well as traceability, excellence and low planning expense. Some of the soft computing were presented for application in medical based arenas a couple of decades ago. The diagnosis framework based on soft computing uses signs for identifiable evidence of syndromes. Certain side effects of the syndrome might be noticed and they are medical parameters like pulse level, glucose level in blood, reports and so on. A computational system is provided by soft computing to tackle formation, study as well as model problems with respect to doubtful and uncertain data [15]. Fuzzy Logic and Neural Network as well as Genetic Algorithm are the Components of soft computing which share a synergetic bond rather than an economical one. In several domains like therapeutic, education, finance, commercial etc, these practices have been related [14].

Artificial intelligence attempts to build computers smarter. Learning is a fundamental necessity for any intelligent activity. Today the majority of the scholars come to an agreement that there is definitely no intelligence deprived of

[1]Research Scholar, Dept of Computer Science,
Karpagam Acadmy of Highzer Education Coimbatore

[2]Asst.Professor, Dept of CS,CA & IT,
Karpagam Academy of Higher Education, Coimbatore

education. So, ML is one of chief topics of AI besides being one of the most quickly emerging subfields of AI research. From the very beginning ML algorithms were also applied to analyze medical data sets. Nowadays ML offers numerous essential means for intellectual data analysis. Specifically, in the past years, moderately cheap as well as available means of the digital revolution was provided to gather and save data [5].Disease-prediction is one of the remarkable as well as challenging responsibilities for doctors. ML methods as a result, have turned out to be a popular tool for medical scholars. These methods can find patterns from complex datasets and connections among them, although they are capable to efficiently predict future outcomes of any disease forms [2].In using classification methods there is an aggregate importance in clinical research. Classification approaches agree to assign topics to one of a mutually complete set of states. The disease conditions (present / absence of disease) or the correct classification of the disease etiology otherwise subtype agree to successive research, actions also involvements to be carried out in a more targeted method which is effective.

Likewise, accurate description of the disease states leads to further correct assessment of patient diagnosis.[4] ML algorithms have been used successfully to create CADx (computer-aided diagnosis) systems. Using the diagnosed samples these algorithms are initially trained, i.e. through standard analyses of medical specialists. The algorithms in the test phase, are subsequently used to support the specialists in the analysis of future samples. Success of an investigation approach in this phase can be described as the capability of algorithm to predict the exact condition (regular or disease) of hidden data [6]. ML is a wide-ranging, quick-developing topic that involves computer applications, statistical data, data mining, as well as optimization technique. The smartness of ML originates from its robust basis in statistical model, which leads to mostly related, certified algorithms in contrast to distinctive medical algorithms that have handcrafted instructions using exemptions of exceptions. The aim of an ML algorithm is to guide ("learn") from present data to identify patterns as well as create knowledgeable decisions [7].Perhaps the most extensive application of ML-built analysis emerging in publications is in neurodegenerative syndromes, in which scholars mean to analyze Alzheimer's or other methods of dementia and expect translation from MCI-mild cognitive impairment to dementia, relying on MR images of the brain. This is probable determined, at least in part, by the accessibility of big datasets with analytical tags, like ADNI-Alzheimer's disease Neuro imaging Initiative as well as OASIS-Open Access Series of Imaging Studies [8].

Reduced cognition in Bipolar Disorder (BD) patients has been stated by multiple Neuro cognition reports as associated with healthy persons. For instance, a new meta-analysis reviewing 42 studies detected substantial injuries in the BD patients within more than one field like attention, functioning memory, visuospatial act, psychomotor speed and philological as well as supervisory function. Numerous studies have effectively spotted distinctive Neuro cognitive subdivisions with data-driven ML algorithms [9]. Recently, ML methods have been established to identify tumor areas in histopathological measures Besides, software made by theoretical or saleable developers proficiently detached cancer from stroma within image blocks on a sub-block resolution. Nonetheless, numerical analysis of samples marked with H&E, which is routinely used for histopathological assessments, is even more complicated, apart from the growth of application software for H&E marked slides being the main factor in the development of the field of image analysis. [10].

Generally, for nearly two decades, ML-based methods have also been applied to research in engineering complications. This is where the use of these methods in both the geosciences and the remote sensing field is honestly new and also very few [11].Numerous ML algorithms like NN-neural networks, linear regression, KNN- k-nearest neighbours, random forest, as well as SVM-support vector machine, shows that these have high potential for modeling relationship between WSS and parameters of geometrically

parameterized models of abdominal aortic aneurysm (AAA) and carotid bifurcation and these algorithms are all unstructured [12]. Until now, in addition to high performance sequence approaches, there is an overabundance of digital systems as well as sensors from many research areas creating data, comprising super resolution digital microscopy, mass spectrometry, MRI, etc. Even though these techniques yield an abundance of data, they don't offer any sort of analysis, explanation or knowledge. The field of Biological Data Mining or the extraction of information in Biological Data is therefore beyond requirement and significance. In addition to applying ML and data mining approaches in Diabetes mellitus (DM) research is a main method for using a huge volume of accessible diabetes-based data for extracting knowledge. ML and data mining approaches in DM are of course highly apprehensive when it comes to diagnosis and management as well as additional associated clinical administration functionalities [3].

## 2. LITERATURE SURVEY

M Chen. 2017[16] rationalized ML algorithms intended for successfully predicting chronic disease occurrence in disease-recurrent societies. The adapted prediction patterns were tested over real hospital data obtained from some parts of China during 2013 to 2015. To deal with the complications of lacking data, a new factor model was used to recreate the missing data. The experiment was carried out on a chronic regional cerebral infarction condition. Based on multimodal disease risk prediction algorithm a novel convolution neural network (CNN)was proposed using structured and unstructured hospital data. Based on the knowledge gained, no current work focused on either type of data in the area of Big Data Clinical Analytics. Matched with a number of typical algorithms for prediction, the prediction rate of the algorithm recommendedachieved94.8% with a convergence speed, and enhanced CNN-based uni modal algorithm for disease risk prediction.

M Nilashi: 2018[17]considered the benefits of an incremental ML technique, Incremental SVM, using a novel process for UPDRS-Unified Parkinson's Disease Rating Scale prediction. To predict Total-UPDRS and Motor-UPDRS the Incremental SVM was used. Also, for data dimensionality reduction a Non linear iterative partial least square was used for clustering a self organizing map. Many experiments led to evaluate the scheme with a Parkinson's disease (PD) dataset, and then presented the results in contrast with the earlier research methods. The outcomes of the experimental study revealed that the recommended method was successful in UPDRS prediction. The scheme had the capability to be applied as an intellectual system for prediction of PD in healthcare.

G Manogaran et al. 2017[18] used Clustering approach with the combination of a Bayesian hidden Markov model (HMM) with Gaussian Mixture (GM) to exhibit the DNA copy number variation across the genome. With many prevailing approaches the proposed Clustering approach was related to those such as Pruned Exact Linear Time method, binary segmentation approach in addition to segment neighbourhood approach. The proposed change detection algorithm's effectiveness is shown in the Experimental results.

DH Alonso et al. 2018[19] used ML based on the combination of adenosine myocardial perfusion SPECT (MPS) findings besides new clinical variables, which predicted a patient's risk of cardiac death, which improved the prediction accuracy and also reduced the input variable numbers. Also have exposed that the SVM algorithm possibly will be the most reliable algorithm but it needs huge volume of features in addition to a minimum absolute shrinkage besides selection operator (LASSO) model and might achieve good accuracy through only six variables. To let the clinician understand the results, a tool for data-visualization is determined on the rationale behind the LASSO's risk score. The percentile rank was also presented appreciating the relative importance of the characteristics of the patient.

MR Mohebian *et al*. 2017[20] presented a scheme for prediction of 5-year old breast cancer reappearance. 579

breast cancer patients Clinical pathological features were studied; the discriminative characteristics were also analyzed using statistical character selection methods. Using PSO- Particle Swarm Optimization they were processed, as input of classification systems with learning ensemble BDT-Bagged Decision Tree. The PSO algorithm identifies the blend of certain categorical features as well as the weight(status) of the particular interval-measurement scale features. Using the holdout and 4-fold cross-validation the performance result of Hybrid Predictor of Breast Cancer Recurrence was measured. It exhibited outstanding agreement using the gold standard (i.e. the opinion of the oncologist after blood tumor marker as well as imaging tests, also tissue biopsy). This algorithm therefore is proved to be a favorable breast cancer reappearance prediction tool.

## 3. CONCLUSION

The development of machine learning and its application in prediction of diseases shows that algorithms, systems and methodologies have emerged from simple and straightforward use that allow for advanced and sophisticated data analysis. In the future, intelligent data processing will play an even greater role, given the enormous amount of knowledge generated and processed by modern technology. Machine learning algorithms currently have tools that can help medical practitioners discover important relationships in their data in a significant way.

## 4. REFERENCES

[1] Nilashi, Mehrbakhsh, Othman bin Ibrahim, Hossein Ahmadi, and Leila Shahmoradi. "An analytical method for diseases prediction using machine learning techniques." Computers & Chemical Engineering 106 (2017): 212-223.

[2] Kourou, Konstantina, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, and Dimitrios I. Fotiadis. "Machine learning applications in cancer prognosis and prediction." Computational and structural biotechnology journal 13 (2015): 8-17.

[3] Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. "Machine learning and data mining methods in diabetes research." Computational and structural biotechnology journal15 (2017): 104-116.

[4] Austin, Peter C., Jack V. Tu, Jennifer E. Ho, Daniel Levy, and Douglas S. Lee. "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes." Journal of clinical epidemiology 66, no. 4 (2013): 398-407.

[5] Kononenko, Igor. "Machine learning for medical diagnosis: history, state of the art and perspective." Artificial Intelligence in medicine 23, no. 1 (2001): 89-109.

[6] Ozcift, Akin, and Arif Gulten. "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms." Computer methods and programs in biomedicine 104, no. 3 (2011): 443-451.

[7] Kang, John, Russell Schwartz, John Flickinger, and Sushil Beriwal. "Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective." International Journal of Radiation Oncology* Biology* Physics 93, no. 5 (2015): 1127-1135.

[8] de Bruijne, Marleen. "Machine learning approaches in medical image analysis: From detection to diagnosis." (2016): 94-97.

[9] Wu, Mon-Ju, Benson Mwangi, Isabelle E. Bauer, Ives C. Passos, Marsal Sanches, Giovana B. Zunta-Soares, Thomas D. Meyer, Khader M. Hasan, and Jair C. Soares. "Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning." Neuroimage 145 (2017): 254-264.

[10] Gertych, Arkadiusz, Nathan Ing, Zhaoxuan Ma, Thomas J. Fuchs, Sadri Salman, Sambit Mohanty, Sanica

Bhele, Adriana Velásquez-Vacca, Mahul B. Amin, and Beatrice S. Knudsen. "Machine learning approaches to analyze histological images of tissues from radical prostatectomies." Computerized Medical Imaging and Graphics 46 (2015): 197-208.

[11] Lary, David J., Amir H. Alavi, Amir H. Gandomi, and Annette L. Walker. "Machine learning in geosciences and remote sensing." Geoscience Frontiers 7, no. 1 (2016): 3-10.

[12]Jordanski, Milos, Milos Radovic, Zarko Milosevic, Nenad Filipovic, and Zoran Obradovic. "Machine learning approach for predicting wall shear distribution for abdominal aortic aneurysm and carotid bifurcation models." IEEE journal of biomedical and health informatics 22, no. 2 (2018): 537-544.

[13] Ma, Chuang, Hao Helen Zhang, and Xiangfeng Wang. "Machine learning for big data analytics in plants." Trends in plant science 19, no. 12 (2014): 798-808.

[14] Long, Nguyen Cong, Phayung Meesad, and Herwig Unger. "A highly accurate firefly based algorithm for heart disease prediction." Expert Systems with Applications 42, no. 21 (2015): 8221-8231.

[15] Shanmugam, S., J. Preethi, and Tamil Nadu. Study of Early Prediction and Classification of Arthritis Disease using Soft Computing Techniques. Infinite Study.

[16]Chen, Min, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. "Disease prediction by machine learning over big data from healthcare communities." IEEE Access 5 (2017): 8869-8879.

[17] Nilashi, Mehrbakhsh, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi, and Mohammadreza Farahmand. "A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques." Biocybernetics and Biomedical Engineering 38, no. 1 (2018): 1-15.

[18] Manogaran, Gunasekaran, V. Vijayakumar, R. Varatharajan, Priyan Malarvizhi Kumar, Revathi Sundarasekar, and Ching-Hsien Hsu. "Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering." Wireless personal communications(2017): 1-18.

[19] Kumar, Priyan Malarvizhi, and Usha Devi Gandhi. "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases." Computers & Electrical Engineering 65 (2018): 222-235.

[20] Alonso, David Haro, Miles N. Wernick, Yongyi Yang, Guido Germano, Daniel S. Berman, and Piotr Slomka. "Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning." Journal of Nuclear Cardiology(2018): 1-9.

[21] Mohebian, Mohammad R., Hamid R. Marateb, Marjan Mansourian, Miguel Angel Mañanas, and Fariborz Mokarian. "A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning." Computational and structural biotechnology journal15 (2017): 75-85.