A Survey on Diverse Chronic Disease Prediction Models and Implementation Techniques

Nitin Chopde and Rohit Miri

Abstract In today's world, chronic diseases are a crucial reason for death. The chronic disease is gradually taking the patient into control and then take over. Chronic start slowly and continues for a long time. There is a need to predict chronic disease at early stages before it reaches an uncontrolled situation so timely treatments can resist it. Prediction system effectively controls chronic disease at early stages. Our study aims to cover various prediction models for chronic disease and techniques for the development of prediction models. This review gives a comprehensive overview of the predictions system and implemented techniques for basic chronic disease. We go through prediction models are developed for basic chronic diseases like heart disease, cancer, diabetes and kidney disease with a different set of techniques. The survey paper discusses an overview of different chronic disease prediction models and its implementation techniques. The survey shows that machine learning approach is efficient to design a prediction system for chronic diseases in the welfare of health organizations and ultimate benefit to patients. This paper reviews basic chronic disease prediction models and suggested that to achieve accurate results of chronic disease prediction system machine learning is promising.

Keywords Chronic disease · Prediction model · Classifier

1 Introduction

They proposed hybrid method and filtering method classifications for a chronic system. The decision tree algorithms based on classification are used to develop a predictive model that predicts risk diagnosis using the observed things [1]. This paper starts from handling extreme values to the use of predictive analytics for selecting

N. Chopde (✉) · R. Miri
Department of Computer Science & Engineering, Dr. C. V. Raman University Kota, Bilaspur (C.G.), India
e-mail: nitin.rchopde@gmail.com

R. Miri
e-mail: rohitmiri@gmail.com

S. C. Satapathy et al. (eds.), *Intelligent System Design*, Advances in Intelligent Systems and Computing 1171, https://doi.org/10.1007/978-981-15-5400-1_51

optimal subset of parameters. It selects the best parameter for prediction and applied strategy to begin from acute values for the prediction [2]. To predict disease they studied genetic issues of polygenic diseases [3] Author forced on accuracy for chronic kidney disease dataset in which various parameters processed using two critical [4]. They studied multiple machine learning algorithms to predict the risk of chronic disease diabetes disease, heart disease, hepatitis disease and liver disease; also they evaluate machine learning algorithms and tools for decision making and prediction of disease [5].

They concentrate on claim data in which they studied clinical and claim data of chronic disease arthritis, kidney disease, osteoporosis [6]. Authors proposed neural network models for CKD detection may effectively and feasibly equip medical staff with the ability to make precise diagnosis and treatment to the patient [7]. They developed neural network models to diagnose chronic kidney disease and treatment to patient which improve the accuracy of prediction. To improve the performance in terms of accuracy of chronic disease prediction they applied adaptive and parallel classification systems [8]. Machine learning effectively classifies healthcare data, as well as diagnosis efficiently and generates accurate results [9].

2 Literature Survey

To provide security to handle big data of the health industry, they introduced a novel model using advanced security techniques along with machine learning algorithms. The innovation lies in the incorporation of optimal storage and data security layer used to maintain security and privacy [10]. They performed comparative analysis of different machine learning algorithms and suggested that ANN is good for prediction but it is slow as compared to other algorithms. Trees algorithm useful for prediction but its complexity is more and to maximum output, you should use RS theory [5].

To get deep information about Medicare data, codebook from CMS is available [11]. In [12] proposed model they used different classification algorithm of machine learning like K-Nearest Neighbour (KNN), tree-based decision tree, LDA classifier and lazy learner for decision support system also they analyzed probability-based Naive Bayes neural network based back propagation (BPN) [13]. These algorithms have been compared with classification accuracy to each other on the basis of correctly classified instances, mean absolute error, Kappa statistics and RMSE metric. The results show that MLP classifier outperforms Naïve Bayes classifier in all sectors with respect to the parameters specified [14].

Then Machine Learning techniques were applied to derive the correlation between the 11 chronic diseases and the ICD9 diagnostic codes. Reduced set of diagnostic codes were obtained for each chronic disease in the training phase. These codes were tested on a test data sample using accuracy and confusion matrix as the performance metric. It was understood that these reduced set of diagnostic codes are relevant and important for each chronic disease [12]. They proposed accuracy point of view for extraction rule best choice is SQR ex-SVM and eclectic approaches for SVM

and experimental result shows that the best approach is C5 [15]. They investigate prediction by machine learning over big data which comes from healthcare and archives accurate result [16].

They proposed chronic disease prediction model, where they evaluated different algorithms like decision tree (J48) and RF algorithm ensemble classifier on various parameters like ROC performance measures, accuracy, precision, recall and F-measure and experimental results both algorithm performance is superior [17].

Authors constructed a system which used UCI repository dataset having 400 data records and 25 attributes and concluded highest accuracy 99.1% from Multiclass Decision forest algorithm [18]. The author implemented AdaBoost machine learning ensemble technique for diabetic diagnostic and result shows in line with J48 and decision tree it performs well [19]. They designed a system for diabetes prediction, focused on the diagnosis of the patient at a particular age; to achieve this, they have implemented a system using machine learning algorithm decision tree they found higher accuracy with decision tree [20, 21]. The author applied classification with a rule for mining for healthcare data [22]. The author has experimented prediction system for diabetic patients in two parts—one is artificial neural network and second one is fast blood sugar [23]. The author classifies biological data into different classes for this they applied rule-based classifier due to this it filters out many problems. And provide a solution for noisy instances and over fitting problems [24].

3 Analysis of Various Approaches and Techniques for Prediction Model

Decision tree classifications applied to predict chronic disease risk diagnosis [1]. The authors have developed prediction model using advances in machine learning to recognize intelligent solutions that enhance prediction in chronic kidney disease [2].

They investigate K-Nearest Neighbour, J48 tree Bayesian Network and Random tree in context with error rate, accuracy and learning time and suggested that on classification Bayesian algorithm is healthy. They had compared the analysis of classification function techniques for heart disease prediction [1]. The heart disease prediction is done by using classification and applied Sequential Minimal Optimization algorithm, LMT algorithm, Multilayer Perception algorithm [4].

This paper focused on clinical and claims data for studying 11 chronic diseases such as kidney disease, osteoporosis, arthritis, etc. using the claims data. The correlation between chronic diseases and the corresponding diagnostic tests is analyzed, by using ML techniques [6]. Hadoop/Map Reduce techniques was used in which predictive analysis algorithm predict the diabetes types prevalent, complications associated with it and further analysis, this system provides an efficient way to cure and care the patients [25]. They used a special model SVM named black box model which gives

beneficial information for SVM's pinpointing decision. SVM is the best solution for diagnosing real-life diabetic problems [15].

Authors predict kidney diseases by applying basic data mining techniques such as clustering, regression, time series and sequence analyses, classification, association analysis, and if the proper technique of data mining is applied it gives promising results [26]. Authors developed a prediction model that diagnosed chronic kidney failure disease using c4.5 learning algorithm which is machine learning algorithms astutely exploit information and extract the best knowledge. Authors designed predictive model for Egyptian patients' using Artificial Neural Networks is machine learning approach where responses are based on their clinical and biochemical data [27] (Table 1).

4 Conclusion

In this survey, it is found that machine learning gives promising accuracy of prediction model of chronic disease as compared to other techniques like data mining. This paper gives an analysis of various machine learning techniques for different chronic diseases; we covered chronic disease as diabetes disease, kidney disease, heart disease and hepatitis disease. Many authors have experimented with different sets of machine learning algorithm which result in an acceptable level of accuracy for healthcare.

Many efforts have been taken by researchers to develop an efficient prediction system for chronic diseases and researcher used different sets of data mining techniques and machine learning techniques. The study previous research shows that for kidney disease ensemble Random forest classifier gives 100% accuracy with less set of attribute and maximum instances from given dataset and J48 decision tree gives an accuracy of 99%. In the case of diabetic disease, SVM gives 94% accuracy and 93% sensitivity. For heart disease SVM provides good accuracy rate up to 95%. In case of chronic Hepatitis Neural network with back propagation shows highest accuracy of 98% although it takes more time to show maximum output. In case of heart disease SVM gives maximum accuracy with 95.2%.

We contributed by reviewing various recent and past research of chronic disease prediction system and it is suggested that machine learning is a promising approach to design efficient and effective prediction systems in which researcher can filter out various classifiers as per the need of prediction system of chronic disease and achieve a promising result.

Table 1 Comparative analysis of basic chronic disease prediction models with its implemented technique

Types of chronic disease	Implemented techniques and algorithm for prediction model	Outcome
Heart disease and kidney disease [28]	k-nearest neighbours	Standard accuracy of 90%
Heart disease and diabetic [11]	Naive Bayes and support vector machine (SVM)	SVM gives highest accuracy rate of 95.56% as compared to Naïve Bayes classifier has 73.58%
Heart disease [9]	Data mining techniques Naïve Bayes followed by neural network and decision trees	Neural network gives more accurate predictions (49.34%) Naïve Bayes (47.58%) and decision trees (41.85)
Heart disease [27]	Data mining, support vector machine (SVM) and ANN	SVM accuracy 95.2% and artificial neural network accuracy 94.27%
Diabetic disease [29]	Weka tool used along with Naïve Bayes, support vector machine, and functional trees	Support vector machine gives 88.3% accuracy
Diabetic [25]	predictive analysis algorithm in Hadoop/MapReduce	It is an efficient way to cure and care the patients with affordability and availability
Diabetes [15]	Data mining and machine learning approach support vector machines	Accuracy is 94%, sensitivity of 93%, and specificity of 94%
Diabetic [27]	Data mining techniques Naive Bias, SVM, decision tree and artificial neural network	Highest accuracy decision tree 86.47% and SVM 87.32%
Pima Indians diabetes database [30]	Tree, SVM and Naïve Bayes	Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms
kidney [26]	Naive Bayes(NB), J48, and random forest (RF), bagging, AdaBoost	J48 decision tree 99% and random forest an average accuracy 100%
Chronic kidney failure and heart disease [29]	k-nearest neighbours	Accuracy of 90% with error rate 5%
Chronic kidney disease [31]	Multiclass decision jungle, forest, multiclass neural network and logistic regression	Multiclass decision forest algorithm gives 99.17% accuracy
Kidney disease [32]	Naïve base classifier and ANN	Naïve base classifier gives 100% accuracy and ANN gives 72.73%
Chronic kidney disease [17]	Decision tree, J48 and SMO classifier and ensemble classifier random forest	Random forest gives 100% accuracy

(continued)

Table 1 (continued)

Types of chronic disease	Implemented techniques and algorithm for prediction model	Outcome
Chronic Hepatitis [27]	Machine learning algorithm Artificial Neural Networks (ANN) and decision trees (DT)	ANN and prediction accuracy 0.76 and 0.80
Chronic Hepatitis [33]	Data mining algorithms	Neural network with back propagation shows highest accuracy of 98%
Chronic Hepatitis [33]	Data mining techniques CART ID3 Algorithm C4.5 algorithm and binary decision tree algorithm	CART algorithm accuracy of 83.2% and C4.5 gives 71.14%

References

1. Hussein, A. S., Omar, W. M., & Xue, L. (2012). Efficient chronic disease diagnosis prediction and recommendation system. In *Biomedical Engineering and Sciences (IECBES), IEEE EMBS Conference* (pp. 209–214).
2. Aljaaf, A. J., Al-Jumeily, D., Haglan, H. M., Alloghani, M., Baker, T., Hussain, A. J., & Mustafina, J. (2018). Early prediction of chronic kidney disease using machine learning supported by predictive analytics. In *IEEE Congress on Evolutionary Computation (CEC)* (pp. 1–9).
3. Grasso, M. A., Dalvi, D., Das, S., & Gately, M. (2011). Genetic information for chronic disease prediction. In *IEEE International Conference on Bioinformatics and Biomedicine Workshops* (p. 997).
4. Jena, L., & Swain, R. (2017). Work-in-progress: Chronic disease risk prediction using distributed machine learning classifiers. In *International Conference on Information Technology (IEEE).*
5. Fatima, M., & Pasha, M. (2017). Survey of Machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications, Scientific Research Publishing* (pp. 1–16).
6. Gupta, D., Khare, S., & Aggarwal, A. (2016). A method to predict diagnostic codes for chronic diseases using machine learning techniques. In *International Conference on Computing, Communication and Automation (ICCCA2016)* (pp. 281–287). IEEE.
7. Kei Chiu, R., Chen, R. Y., Wang, S., & Jian, S. (2012). Intelligent systems on the cloud for the early detection of chronic kidney disease. In *Machine learning and cybernetics* (pp. 1737–1742). IEEE.
8. Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction. *Egyptian Informatics Journal, 19,* 179–189.
9. Tayeb, S., Pirouz, M., Sun, J., Hall, K., Chang, A., Li, J., et al. (2017). Toward predicting medical conditions using k-nearest neighbors. In *IEEE International Conference on Big Data (BIGDATA)* (pp. 3897–3903).
10. Kaur, P., Sharma, M., & Mittal, M. (2018). Big data and machine learning based secure healthcare framework. In *International Conference on Computational Intelligence and Data Science (ICCIDS)* (pp. 1049–1059). Elsevier.
11. Palaniappan, S. P., & Awang, R. (2008). Intelligent heart prediction system using data mining techniques, IEEE (pp. 108–115).
12. Ruiz-Arenas, R. (2017). A summary of worldwide national activities in chronic kidney disease (CKD) testing. *The Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, 28*(4), 302–314.

13. Ani, R., Sasi, G., Sankar, U. R., & Deepa, O. S. (2016). Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification. In *Advances in Computing, Communications and Informatics (ICACCI), International Conference IEEE* (pp. 1287–1292).
14. Lin, R. H., Ye, Z. Z., Wang, H., & WuI, B. (2018). Chronic diseases and health monitoring big data: A survey. *IEEE Transactions Journal*, 1 –15.
15. Mellitus, Barakat, N. H., & Bradley, A. P. (2010). Intelligible support vector machines for diagnosis of diabetes senior member. *IEEE Transactions on Information Technology in Biomedicine, 14*(4).
16. Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 1–9.
17. Sisodia, D. S., & Verma, A. (2017). Prediction performance of individual and ensemble learners for chronic kidney disease. In *Proceedings of the International Conference on Inventive Computing and Informatics IEEE Xplore* (pp. 1027–1031).
18. Gunarathne W. H. S. D, Perera, K. D. M, & Kahandawaarachchi, K. A. D. C. P. (2017). Performance evaluation on machine learning classification techniques for classification and forecasting through data analytics for chronic kidney disease. In *IEEE International Conference on Bioinformatics and Bioengineering.*
19. Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science.*
20. Orabi, K. M., Kamal, Y. M., & Rabah, T. M. (2016). Early predictive system for diabetes mellitus disease. In *Industrial Conference on Data Mining* (pp. 420–427).Springer.
21. Rashid, T. A., & Abdullah, R. M. (2016). An intelligent approach for diabetes classification, prediction and description. *Advances in Intelligent Systems and Computing*, 323–335.
22. Song, S., Warren, J., & Riddle, P. (2014). Developing high risk clusters for chronic disease events with classification association rule mining. In *CRPIT—Health Informatics and Knowledge Management* (Vol. 153, pp. 69–78).
23. NaiArun, N., & Moungmai, R. (2015). Comparison of classifiers for the risk of diabetes prediction. *Procedia Computer Science, 69*, 132–142.
24. Farid, D. M., Al-Mamun, M. A., Manderick, B., & Nowe, A. (2016). An adaptive rule-based classifier for mining big biological data. *International Journal of Expert Systems with Applications, 64*, 305–316.
25. Saravana Kumar, N. M., Eswari, T., Sampath, P., & Lavanya (2015). Predictive methodology for diabetic data analysis in big data. In *2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).*
26. ElHefnawi, M., Abdalla, M., Ahmed, S., Elakel, W., Esmat, G., Elraziky, M., et al. (2012). Accurate prediction of response to Interferon-based therapy in Egyptian patients with chronic hepatitis C using machine-learning approaches. In I*EEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 772–778).
27. Deepika, K., & Seem, S. (2016). Predictive analytics to prevent and control chronic diseases. In *IEEE 2nd International Conference on App Applied and Theoretical Computing and Communication Technology* (pp. 381–386).
28. Wickramasinghe, M. P. N. M., Perera, D. M., & Kahandawaarachchi, K. A. D. C. P. (2017). Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms (pp. 300–303). IEEE.
29. Suresh Kumar, P., & Pranavi, S. (2017). Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In *International Conference on Infocom Technologies and Unmanned Systems* (pp. 508–513).
30. Sisodiaa, D., & Sisodiab, D. S. (2018). Prediction of diabetes using classification algorithms. In *International Conference on Computational Intelligence and Data Science (ICCIDS 2018.*

31. Dilli Arasu, S. (2017). Review of chronic kidney disease based on data mining techniques. *International Journal of Applied Engineering Research, 12*(23), 13498–13505.
32. Jena, L., & Kamila, N. K. (2015). Distributed data mining classification algorithms for prediction of chronic kidney-disease. *International Journal of Emerging Research in Management & Technology*, *4*(11). ISSN: 2278–9359.
33. Sathyadevi, G. (2011). Application of CART algorithm in hepatitis disease diagnosis. In *IEEE International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 1283–1287).