International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 4 Issue 11, November 2023

Chronic Diseases Classification through Machine Learning Algorithms: A Survey

Jyotika Shrivastava[1], Chetan Agrawal[2], Rashi Yadav[3]
Dept. of CSE, Radharaman Institute of Technology & Science, Bhopal, India[1,2,3]
jyotikashrivastava377@gmail.com[1], chetan.agrawal12@gmail.com[2], rashi6yadav@gmail.com[3]

***Abstract**: Chronic diseases pose a significant global health burden, with early detection and diagnosis crucial for improved patient outcomes. Traditional diagnostic methods often rely on subjective clinical assessments and invasive procedures, limiting their accessibility and timeliness. Machine learning (ML) algorithms offer a promising approach to chronic disease classification, enabling the analysis of large datasets of patient data to identify patterns that aid in early disease detection. This survey delves into the application of ML algorithms for chronic disease classification, exploring the diverse techniques employed by researchers and their associated accuracy levels. The role of feature selection and independent variable selection in enhancing algorithm performance is emphasized, highlighting the importance of selecting the most relevant features from the data. Additionally, the study underscores the benefits of combining multiple algorithms to achieve superior accuracy compared to single algorithms. The development of disease prediction systems using ML holds the potential to revolutionize healthcare by enabling symptom-based diagnoses. Selecting the most appropriate ML model is paramount for making informed decisions regarding chronic disease diagnosis.*

Keywords: Chronic diseases, machine learning, disease classification, early detection, diagnosis, Healthcare revolution, feature selection

1. Introduction

All over the world, chronic diseases are a critical issue in the healthcare domain. According to the medical statement, due to chronic diseases, the death rate of humans increases. The treatments given for this disease consume over 70% of the patient's income. Hence, it is highly essential to minimize the patient's risk factor that leads to death. The advancement in medical research makes health-related data collection easier [1, 2]. The healthcare data includes the demographics, medical analysis reports, and the history of disease of the patient. The diseases caused could be varied based on the regions and the living habitats in that region. Hence, along with the disease data, the environmental condition and the living habitat of the patient should also be recorded in the data set.

In recent years, the healthcare domain is evolving more due to the integration of information technology (IT) in it. The intention to integrate IT in healthcare is to make the life of an individual more affordable with comfort as smart- phones made one's life easier [3]. This could be possible by making healthcare to be intelligent, for instance, the invention of the smart ambulance, smart hospital facilities, and so on, which helps the patients and doctors in several ways [4]. The research on a specified region for patients affected by chronic diseases every year had been held and found that the difference between the patients in gender wise is very small, and it is found that the large number of patients were ad- mitted in the year 2014 for treating chronic diseases. The use of structured and unstructured data provides highly accurate results instead of using only structured data. Since the unstructured data includes the doctor's records on the patients related to diseases and the patient's symptoms and grievances faced by them, explained by themselves, which is an added advantage when used along with the structured data that consists of the patient demographics, disease de- tails, living habitats, and laboratory test results [5, 6]. It is difficult to diagnose rare diseases. Hence, the use of self-reported behavioral data helps differentiate the individuals with rare diseases from the ones with common chronic diseases. By using machine learning approaches along with questionnaires, it is believed that the identification of rare diseases is highly possible [7]. Figure 1 shows the conceptual view of the CKD prediction model in Machine Learning.

International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 4 Issue 11, November 2023

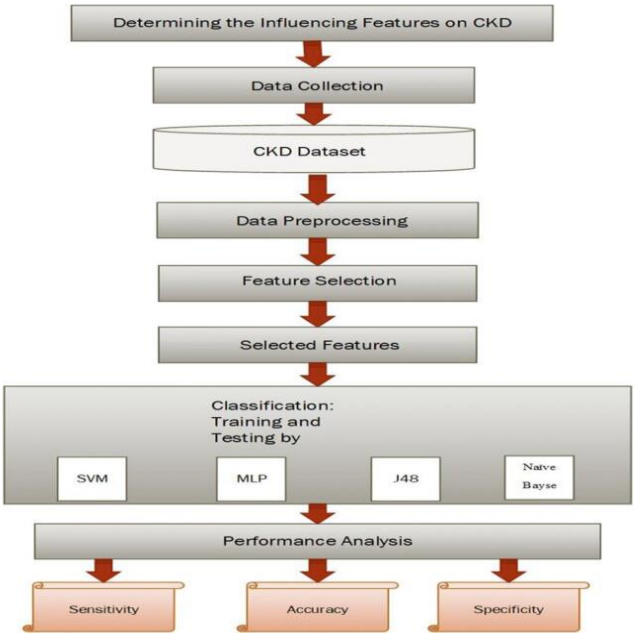

Figure 1: The conceptual graph CKD prediction model using Machine Learning Method

In the last decade, some innovative technologies had been introduced to rapidly collect the data such as MRI (magnetic resonant imaging) readouts, ultra-sonography, social media gained data, and electronically gained activity, behavioral, and clinical data. These big data sets of healthcare are high-dimensional, which means the number of features recorder per observation might be greater than the total observations. They are noisy, sparse, cross-sectional, and lacks statistical power. By applying machine learning techniques, the issues in the high-dimensional data sets can be overcome [8]. Machine learning contributes more in several domains. Many of the complex models make use of exiting larger training data, simultaneously at the edge of a major shift in healthcare epidemiology [9]. These data can enhance the knowledge gain in the risk factors of diseases to reduce healthcare-associated infections, improve patient risk stratification, and find the way of transmitting the infectious diseases [10]. Machine learning can facilitate the analysis of laboratory results and other details of patients for the early detection of diseases. The low-level data could be converted to high-level knowledge via knowledge discovery in the database so as to gain knowledge about the disease patterns to support early detection [11]. The data collected for creating a data set should be preprocessed for its missing values, and then only the important features needed for accurate disease prediction are selected so as to enhance the prediction accuracy and minimizing the time taken for model training [12].

In the era of the Internet and technologies, people are not concerned about their health and lives. As everyone is interested in surfing and social media activities, they ignore visiting hospitals for their health checkup. By taking this activity as an advantage, a machine learning model that takes the symptoms given as input and predicts the possibility and risk of the disease affected or the development of such diseases in an individual should be developed [13, 14]. The more common chronic diseases are diabetes, cardiovascular diseases, cancer, strokes, hepatitis C, and arthritis. As these diseases persist for a long time and have a high mortality rate, the diagnosis of such diseases is highly important in the healthcare domain. Foreseeing the disease can help take preventive actions and avoid getting affected by it and early detection of it can help provide better treatment [15]. There are various techniques in machine learning such as supervised, semi-supervised, unsupervised, reinforcement, evolutionary, and deep learning. The problem is associated with the processing of extracted features from real data and structured as vectors [16]. The processing quality is based on the proper combination of those vectors. But, most of the times, the high dimensionality of the vectors or the discrepancies in the data makes a big issue. Hence, it is important to reduce the dimensionality of the data set even if it leads to a little loss of details to make the data set a highly compatible dimension. This reduction in the dimensionality of the data set improves the model performance [17].

The system of chronic diseases management is essential for those affected by such diseases and in need of proper medical assessment and treatment information [18]. Also, this system can be useful for individuals who are in need of self-care to improve their health condition, since it is proved that self-management is the primary care of those with chronic diseases, and it is considered as the unavoidable part of treatment. With the use of mobile applications, the health information of patients can be recorded, and they serve as a better tool to enable self-management [19]. To effectively predict a disease, information such as narration about the symptoms felt by the patients, details of consultation with medical practitioners, lab examination results, and computed tomography and X-ray images [20]. Little research is performed in identifying the accuracy and predictive power for developing a machine learning model with only information from lab examination results for the diagnosis of diseases. And, for performance enhancement, ensemble machine learning and deep learning model can be used [21, 22]. In the healthcare domain, artificial intelligence (AI) plays a major role in automating the roles involved in disease diagnosis and treatment suggestions and also

schedules perfect timing by the medical practitioners to perform various obligations that cannot be automated [23].
In recent decades, chronic disease is a long-lasting illness that has a huge impact on people health. The most frequent chronic disease are hyperlipidemic arthritis, coronary artery diseases, colon cancer, asthma, heart disease, hemophilia, chronic kidney disease, chronic respiratory disease, etc. [4]. The risk behaviours responsible for chronic disease are; hypertension (raised blood pressure), tobacco use, raised cholesterol, unhealthy diet, physical inactivity, and harmful use of alcohol. The risk behaviours responsible for chronic disease is graphically denoted in figure 2.

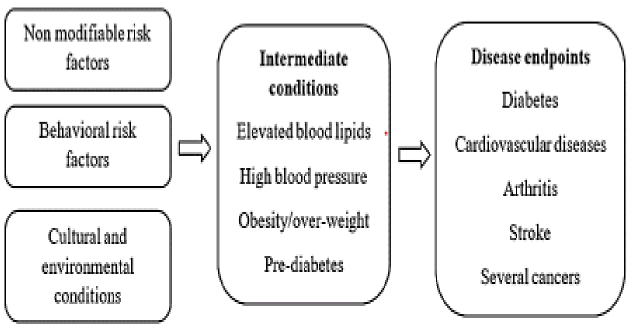

Figure 2: Graphical representation of risk behaviors responsible for chronic disease

The major objective of this system is to identify and predict chronic disease in an individual using a machine learning approach [24, 25]. The data set comprises both the structured data that includes the patient's age, gender, height, weight, and so on, excluding the patient's personal information such as name and ID, and the unstructured data that includes the patient's symptoms, information related to consultation about the disease with the doctors, and the living habits of that individual [26]. These data are pre- processed for finding the missing values. They are then reconstructed to increase the quality of the model, thereby increasing the prediction accuracy. For prediction, the machine learning algorithms such as CNN and KNN are used [27, 28].
This paper is organized as the details of the related works carried out while doing the research are given in Section 2, the problem statement and challenges are given in Section 3, the future research directions are defined in Section 4, followed by the conclusion in Section 5, and finally, a list of references used in this study has been given.

2. Related Work

The main goal of research study is to collect all the published articles related to this field and conclude about the coverage of research done so far. We focused on the published research from 2017 up to today, review the research done on various machine learning techniques and algorithms used for the efficient prediction of diseases in various healthcare applications.
Min Chen et al.[29] are the first to work on both type of data i.e., structured and unstructured and have proposed a multimodal disease risk prediction model based on CNN. In comparison to other predicting algorithms, this model proposed an accuracy of 94.8% with better convergence speed than CNN-UDRP.
Mehrbakhsh et al. [30] bring techniques based on CART(classification and regression trees), EM(Expectation Minimization), PCA(Principal Component Analysis) and fuzzy rule for diagnosis of disease and obtain good accuracy in prediction. Result indicates that the method of combining fuzzy rule-based along with clustering and PCA are successful in obtaining good accuracy.
Dhiraj et al. [31] suggested a model of disease prediction model based on patients symptoms using the CNN and KNN ML algorithms for accurate disease prediction. Disease symptoms dataset is required for disease prediction. In comparison to KNN algorithm, disease prediction accuracy of CNN is much more (84.5%). Along with the prediction of disease, this proposed system was able to predict the risk (low or high) associated with the disease.
A NB based classifier model was proposed by Venkatesh [32] to predict future health status from heart disease data. Result declares that this method provides an accuracy of approximately 97.12%. The primary aim of this research is to predict a patient's future health condition.
For predicting fatty liver disease, Mohaimenul Islam [33] developed a ML algorithm based prediction model. Logistic Regression model performance improved with 76.30% accuracy, 74.10% sensitivity and 64.90% specificity in comparison to other classification techniques. Result reveals that it is possible to use Logistic Regression Model as an important tool for clinical decision making.
Pahulpreet and Shriya [34] applied different classification algorithms for early prediction of disease on three different databases and compare the results. From all the comparisons, accuracy of heart disease detection was 87.1% using Logistic Regression, diabetes was 85.71% using SVM and Breast Cancer detection was 98.57% using AdaBoost classifier.
Sayali and Rashmi [35] proposed a method for predicting whether or not a patient has heart disease using NB and KNN algorithm. The high/low risk of illness was also predicted using CNN-UDRP algorithm. Its only drawback is that it uses only structured data whereas Shraddha [36]

overcomes the limitation of CNN-UDRP and proposed CNN-UDRP algorithm that uses hospital structured and unstructured data and results that accuracy is more and fast. Shweta et al. [37] works on different chronic diseases and applied support vector machine, random forest and decision tree to predict whether or not a patient suffers from disease. Chronicle diseases like diabetes, heart disease and liver disease were included and as a result random forest algorithm worked with higher accuracy.

Chronic disease is a major issue of health worldwide. The main cause of death is chronic diseases as per medical standard report. Hence, to reduce the risk of people's life, Anandajayam [38] works on the analysis of chronicle diseases and uses RNN for structured data and CNN for unstructured data. After this, multiple algorithms like SVM, recurrent neural network, K-Nearest Neighbor, naïve bayes and Decision tree were then processed to examine the accuracy rate of disease risk. As a result, performance of RNN algorithm was better in comparison to other algorithms.

A prediction model was suggested by Induja and Raji [39] to reduce the loss of human lives due to cerebral stroke. For training as well as the testing data, ten-fold cross validation was applied and several classification algorithms such as K-nearest neighbor, Decision tree and Naive Bayes were used to predict the risk of stroke. As a result, Decision Tree shows better performance in comparison to all three with an accuracy of 99%.

Ehtisham [40] used SVM and multilinear regression algorithm to predict diseases and perform testing on multiple algorithms such as CNN, decision tree and k-nearest neighbor etc. In comparison, the combination of SVM and multilinear regression provides higher accuracy in the range of 68%-87%.

In combination with APDFS and HLRM, Sandeep Kumar [41] proposed a framework to improve the predictive accuracy of Chronic Kidney Disease by discovering certain characteristics that are important to the diagnosis of Chronic Kidney Disease. As a result, experimental findings showed that as it identifies the disease with 91.6% accuracy, hence the proposed method is efficient.

Theyazn [42] works on an objective to develop the chronic disease surveillance detection system. The dataset was collected from world-wide resources, include ambiguous objects as well. Hence, to remove the ambiguity from dataset, and to boost the performance of a system, the Rough K-means clustering algorithm was used. Different ML algorithms like NB, SVM, KNN and random forest were compared and achieved the best results for diabetic disease classification with Naïve Bayes and RKM (80.55%), while SVM achieved 100% for kidney disease classification in combination with RKM and SVM achieved 97.53% accuracy metric for cancer disease classification along with RKM.

Algorithm Used	**Accuracy**	**Dataset**	**Disease Predicted**	**Result**
CNN based multimodal disease risk Prediction [29]	94.8%	Real lifeHospital dataset	cerebral infarction	High or lowrisk prediction
CART, EM, PCA with fuzzy rules [30]	-	Public medical data sets taken from	Diabetes and heart disease	Good accuracy inprediction
KNN and CNN [31]	84.5%	Downloaded from UCI ML website	General disease	Accuracy of CNN is More in comparison to KNN
NaiveBayes technique [32]	97.12%	UCI ML repository	Heart disease	Predicts future health condition of patient
RF, SVM, LR, ANN with 10-fold cross validation [33]	76.30%	Data were collected from Taipei Medical University Hospital	Liver Disease	A better result is given by the logistic regression technique.
Classification algorithms – LR, DT, RF, SVM and adaptive boosting [34]	87.1% in Heart Disease detection 85.71% in Diabetes and 98.57% for Breast Cancer detection	UCI Machine Learning Library	Heart disease, Breast cancer, Diabetes	In order to understand the disease risk prediction, this approach predicts low, high and medium risk of heart disease.
CNN-UDRP [35]	65%	UCI Machine Learning Library	Heart disease	Accurate disease risk prediction was achieved.

CNN-MDRP [36]	94.8%	Hospitals data	Heart disease	Prediction of disease is fast and accurate
Random Forest Algorithm [37]	65% 83% 98%	UCI Machine Learning Repository	Liver disease, heart disease and diabetes	Random forest algorithm is more accurate
Recurrent neural network and CNN [38]	97.62%	Online dataset from hospitals	cerebral infarction	RNN works better
K-nearest neighbor, Decision tree and Naïve Bayes [39]	99%	National Stroke Mortality dataset.	cerebral stroke	Decision tree's performance is much better whereas NB classifier's performance was poor.
support vector machine and multilinear regression algorithm [40]	87%	by collecting patient's symptoms and diagnosis from local hospitals and from open source libraries available online	Heart Disease	Multilinear regression algorithm is better in predicting the chance of heart disease
APDFS and HLRM [41]	91.6%	Medical labs and hospitals	Chronic kidney disease	Model performed well in prior prediction of CKD
RKM with NB, SVM, KNN and RF [42]	80.55% 100% 97.53%	ML repository and Kaggle	Diabetes, Kidney and Cancer	Model successfully diagnosed the chronic diseases

3. Problem Statement & Challenges

Chronic diseases are long-lasting health conditions that persist for months or years and pose a significant global health burden. Early detection and diagnosis of chronic diseases are crucial for timely intervention and improved patient outcomes. However, traditional diagnostic methods often rely on subjective clinical assessments and invasive procedures, which can be time-consuming, expensive, and inaccessible.

Machine learning (ML) algorithms have emerged as a promising tool for chronic disease classification, offering the potential to analyze large datasets of patient data and identify patterns that can aid in early disease detection. ML models can be trained on vast amounts of electronic health records, including patient demographics, medical history, laboratory test results, and imaging data, to learn the complex relationships between these features and the presence or absence of chronic diseases.

Challenges:

Despite the promise of ML, developing accurate and reliable chronic disease classification models faces several challenges:

- **Data quality and heterogeneity:** Electronic health records often contain incomplete, inaccurate, or inconsistent data, which can hinder the development of robust ML models.
- **Feature selection and dimensionality reduction:** Extracting meaningful features from large datasets is crucial for ML model performance, but identifying the most relevant features can be challenging.
- **Class imbalance:** Chronic diseases often have a low prevalence, leading to an imbalance in the distribution of disease labels within the dataset. This imbalance can bias ML models towards the majority class, leading to poor performance for minority classes.
- **Model interpretability and Explainability:** Understanding the decision-making process of ML models is essential for gaining trust in their predictions and ensuring they align with clinical expertise.

4. Future Research Directions

Future Research Directions in Chronic Disease Classification with Machine Learning

- **Explainable AI (XAI):** Enhance model interpretability, visualize decision-making processes, and explain complex interactions between features.
- **Multimodal Data Integration**: Combine electronic health records, imaging data, and genetic data to improve model performance, develop multimodal data fusion methods, and address data heterogeneity and synchronization challenges.
- **Personalized Prediction and Risk Stratification**: Develop personalized chronic disease risk prediction models, identify high-risk subgroups for targeted interventions, and inform personalized treatment plans and risk mitigation strategies.
- **Real-time Chronic Disease Monitoring:** Implement ML-powered systems for real-time patient data monitoring, generate timely alerts for healthcare providers, and integrate ML models into wearable devices and telemedicine platforms.
- **Chronic Disease Classification in Resource-Limited Settings:** Develop scalable and adaptable ML models, utilize transfer learning and domain adaptation techniques, and design efficient ML models for limited computational resources and power consumption.
- **Ethical Considerations and Regulatory Compliance:** Address data privacy, bias, and transparency concerns, establish guidelines and frameworks for responsible ML development and deployment, and ensure compliance with regulatory requirements and data protection laws.

5. Conclusion

The area of medical care is witnessing a revolution with the advent of innovative decision support systems powered by machine learning (ML). These models have the potential to transform healthcare by emphasizing patient-centered care, enabling accurate diagnoses, guiding appropriate treatments, and contributing to cost reduction. Researchers have primarily utilized hospital data, categorized into structured and unstructured formats, for developing ML models. While research on both data types is still in its nascent stages, several studies across various domains are exploring disease prediction using both structured and unstructured data. This paper delves into the diverse ML techniques employed by researchers for chronic disease diagnosis. Such disease prediction models hold the promise of diagnosing individuals based on their symptoms. A comprehensive analysis of these techniques reveals the crucial role of feature and independent variable selection in enhancing both accuracy and algorithm performance. Additionally, our findings indicate that combining multiple algorithms outperforms single algorithms in terms of accuracy. Moreover, evaluating multiple sequence combinations of algorithms is essential to identify the most effective combination for chronic disease prediction. The development of disease prediction systems holds the potential to revolutionize healthcare by enabling symptom-based diagnoses. Therefore, selecting the most appropriate model is paramount for making informed decisions regarding chronic disease diagnosis. Looking ahead, advanced AI methods such as deep learning and cognitive computing are poised to play a pivotal role in chronic disease analysis. Incorporating medical reports like MRI scans and X-rays into the datasets can further enhance prediction accuracy.

References

[1] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F Amenta, “Applications of machine learning predictive models in the chronic disease diagnosis,” Journal of Personalized Medicine, vol. 10, no. 2, p. 21, 2020.

[2] B. Manjulatha and P. Suresh, “An ensemble model for pre- dicting chronic diseases using machine learning algorithms,” in Smart Computing Techniques and Applications, pp. 337–345, Springer, New York, NY, USA, 2021.

[3] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, and M.-S. Chen, “Application of classification techniques on development an early-warning system for chronic illnesses,” Expert Systems with Applications, vol. 39, no. 10, pp. 8852– 8858, 2012.

[4] D. Gupta, S. Khare, and A. Aggarwal, “A method to predict diagnostic codes for chronic diseases using machine learning techniques,” in Proceedings of the 2016 International Con- ference on Computing, Communication and Automation (ICCCA), pp. 281–287, IEEE, Greater Noida, India, April 2016.

[5] M. Chen, H. Yixue, K. Hwang, L Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” Ieee Access, vol. 5, pp. 8869–8879, 2017.

[6] R. Ge, R. Zhang, and P Wang, “Prediction of chronic diseases with multi-label neural network,” IEEE Access, vol. 8, pp. 138210–138216, 2020.

[7] H. MacLeod, S. Yang, O. Kim, C. Kay, and S. Natarajan, “Identifying rare diseases from behavioural data: a machine learning approach,” in Proceedings of the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), pp. 130–139, IEEE, Washington, DC, USA, June 2016.

[8] M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste et al., “Applications of machine learning to diagnosis

and treatment of neurodegenerative diseases," Nature Reviews Neurology, vol. 16, no. 8, pp. 440–456, 2020.

[9] I. Preethi and K. Dharmarajan, "Diagnosis of chronic disease in a predictive model using machine learning algorithm," in Proceedings of the 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICST- CEE), pp. 191–96, IEEE, Bengaluru, India, October 2020.

[10] J. Wiens and E. S Shenoy, "Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology," Clinical Infectious Diseases, vol. 66, no. 1, pp. 149–153, 2018.

[11] S. Swaminathan, K. Qirko, T. Smith et al., "A machine learning approach to triaging patients with chronic ob- structive pulmonary disease," PLoS One, vol. 12, no. 11, Article ID e0188532, 2017.

[12] Z. Wang, J. W. Chung, X. Jiang, Y. Cui, M. Wang, and A. Zheng, "Machine learning-based prediction system for chronic kidney disease using associative classification tech- nique," International Journal of Engineering & Technology, vol. 7, pp. 1161–1167, 2018.

[13] A. Kumar and A. Pathak, "A machine learning model for early prediction of multiple diseases to cure lives," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 6, pp. 4013–4023, 2021.

[14] C. Kalaiselvi, "Diagnosing of heart diseases using average K-nearest neighbor algorithm of data mining," in Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 3099–3103, IEEE, New Delhi, India, March 2016.

[15] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: a review," Egyptian Informatics Journal, vol. 19, no. 3, pp. 179–189, 2018.

[16] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," Journal of Applied Science and Technology Trends, vol. 2, no. 1, pp. 10–19, 2021.

[17] V. D. Soni, "Chronic disease detection model using machine learning techniques," International Journal of Scientific & Technology Research, vol. 9, no. 9, pp. 262–266, 2020.

[18] P. Ghosh, S. Azam, A. Karim, M. Jonkman, and M. D. Z. Hasan, "Use of efficient machine learning techniques in the identification of patients with heart diseases," in Pro- ceedings of the 2021 the 5th International Conference on In- formation System and Data Mining, pp. 14–20, Silicon Valley USA, May 2021.

[19] Y. Chang and X. Chen, "Estimation of Chronic Illness Severity Based on Machine Learning Methods," Wireless Communi- cations and Mobile Computing, vol. 2021, Article ID 1999284, 13 pages, 2021.

[20] T. H. H. Aldhyani, A S Alshebami, and M. Y. Alzahrani, "Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms," Journal of Healthcare Engi- neering, vol. 2020, Article ID 4984967, 16 pages, 2020.

[21] H. Fro¨hlich, R. Balling, N. Beerenwinkel et al., "From hype to reality: data science enabling personalized medicine," BMC Medicine, vol. 16, no. 1, pp. 1–15, 2018.

[22] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic diseases diagnosis using machine learning," in Proceedings of the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pp. 1–6, IEEE, Kottayam, India, December 2018.

[23] D. J. Park, M. W. Park, H. Lee, Y.-J. Kim, Y. Kim, and Y. H. Park, "Development of machine learning model for diagnostic disease prediction based on laboratory tests," Scientific Reports, vol. 11, no. 1, pp. 1–11, 2021.

[24] J. Mishra and S. Tarar, "Chronic disease prediction using deep learning," in Proceedings of the International Conference on Advances in Computing and Data Sciences, pp. 201–11, Springer, Valletta, Malta, April 2020.

[25] E. Jeong, S. Osmundson, C. Gao, D. R. V. Edwards, M. Bradley, and Y. Chen, "Learning the impact of acute and chronic diseases on forecasting neonatal encephalopathy," Computer Methods and Programs in Biomedicine, vol. 211, Article ID 106397, 2021.

[26] F. Ceccarelli, M. Sciandrone, C. Perricone et al., "Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models," PLoS One, vol. 12, no. 3, Article ID e0174200, 2017.

[27] Md M. Mottalib, J. C. Jones-Smith, B. Sheridan, and R. Beheshti, "Identifying the risks of chronic diseases using BMI trajectories," 2021, https://arxiv.org/abs/2111.05385.

[28] S. Agarwal, C. Prabha, and M. Gupta, "Chronic diseases prediction using machine learning–A review," Annals of the Romanian Society for Cell Biology, vol. 25, pp. 3495–3511, 2021.

[29] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," IEEE Access, vol. 5, no. c, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[30] M. Nilashi, O. bin Ibrahim, H. Ahmadi, and L. Shahmoradi, "An analytical method for diseases prediction using machine learning techniques," Comput. Chem. Eng., vol. 106, pp. 212–223, 2017, doi: 10.1016/j.compchemeng.2017.06.011.

[31] D. Dahiwade, G. Patle, and E. Meshram, "Designing disease prediction model using machine learning approach," Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019, no. Iccmc, pp. 1211–1215, 2019, doi: 10.1109/ICCMC.2019.8819782.

[32] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, "Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique," J. Med. Syst., vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1398-y.

[33] M. Mohaimenul Islam, C. C. Wu, T. N. Poly, H. C. Yang, and Y. C. Li, "Applications of machine learning in fatty live disease prediction," Stud. Health Technol. Inform., vol. 247, pp. 166–170, 2018, doi: 10.3233/978-1-61499-852-5-166.

[34] P. S. Kohli and A. L. Regression, "2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020," 2020 IEEE 5th Int. Conf. Comput. Commun. Autom. ICCCA 2020, pp. 1–4, 2020.

[35] S. Ambekar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network," Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, pp. 1–5, 2018, doi: 10.1109/ICCUBEA.2018.8697423.

[36] A. Agrawal, H. Agrawal, S. Mittal, and M. Sharma, "Disease Prediction Using Machine Learning," SSRN Electron. J., pp. 6752–6757, 2018, doi: 10.2139/ssrn.3167431.

[37] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic Diseases Diagnosis using Machine Learning," 2018 Int. Conf. Circuits Syst. Digit. Enterp. Technol. ICCSDET 2018, pp. 1– 6, 2018, doi: 10.1109/ICCSDET.2018.8821235.

[38] P. Anandajayam, S. Aravindkumar, P. Arun, and A. Ajith, "Prediction of chronic disease by machine learning," 2019 IEEE Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2019, pp. 1–6, 2019, doi: 10.1109/ICSCAN.2019.8878724.

[39] S. N. Induja and C. G. Raji, "Computational Methods for Predicting Chronic Disease in Healthcare Communities," 2019 Int. Conf. Data Sci. Commun. IconDSC 2019, pp. 1–6, 2019, doi: 10.1109/IconDSC.2019.8817044.

[40] M. E. Farooqui and D. J. Ahmad, "Disease Prediction System using Support Vector Machine and Multilinear Regression," SSRN Electron. J., pp. 331–336, 2020, doi: 10.2139/ssrn.3673232.

[41] S. Hegde and M. R. Mundada, "Early prediction of chronic disease using an efficient machine learning algorithm through adaptive probabilistic divergence based feature selection approach," Int. J. Pervasive Comput. Commun., 2020, doi: 10.1108/IJPCC-04-2020-0018.

[42] T. H. H. Aldhyani, A. S. Alshebami, and M. Y. Alzahrani, "Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms," J. Healthc. Eng., vol. 2020, 2020, doi: 10.1155/2020/4984967.

[43] Naiyer, Vaseem, Jitendra Sheetlani, and Harsh Pratap Singh. "Software Quality Prediction Using Machine Learning Application." Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2. Springer Singapore, 2020.

[44] Pasha, Shaik Imran, and Harsh Pratap Singh. "A Novel Model Proposal Using Association Rule Based Data Mining Techniques for Indian Stock Market Analysis." Annals of the Romanian Society for Cell Biology (2021): 9394-9399.

[45] Md, Abdul Rasool, Harsh Pratap Singh, and K. Nagi Reddy. "Data Mining Approaches to Identify Spontaneous Homeopathic Syndrome Treatment." Annals of the Romanian Society for Cell Biology (2021): 3275-3286.