Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.3124-3130 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Predicting Chronic Diseases and Measuring Drug Effectiveness in Diabetes using Machine Learning and Statistical Methods to aid Clinical Decision Making

[1]Anil B C, [2]Alpna Sharma, [3]Surya Pratap Singh, [4]Dr. P Joel Josephson,

[1] Associate Professor & Head, Department of CSE(AI & ML) JSS Academy of Technical Education, Karnataka, India anil.bc2@gmail.com
[2]Assistant Professor, Vivekananda Institute of Professional Studies- Technical Campus, Delhi, India,alpna@vips.edu
[3]Research Scientist-1, Uttar Pradesh University of Medical Sciences, U.P., India, spsinghadarsh@gmail.com
[4]Associate Professor, ECE, Malla Reddy Engineering College, Telangana, India, joeljosephsonp@gmail.com

How to cite this article: Anil B C, Alpna Sharma, Surya Pratap Singh, P Joel Josephson, (2024). Predicting Chronic Diseases and Measuring Drug Effectiveness in Diabetes using Machine Learning and Statistical Methods to aid Clinical Decision Making. *Library Progress International*, 44(3), 3124-3130.

Abstract
An increase in digitalization of patient health records has generated extensive data within the healthcare sector, yet the complexity and fragmentation of this data have hindered effective utilization for improved patient outcomes. This study examines the use of artificially intelligent systems (AI), the use of machine learning (ML), and big data analysis for medical purposes, with a focus on regulating and prediction of chronic illnesses. Transformation brought by AI and ML in diagnostics, patient care, and predictive analytics, emphasizing the significant role these technologies play in enhancing decision-making, improving data quality, reducing healthcare costs, and facilitating early disease detection and treatment is also discussed. Through a comprehensive review of various studies, the effectiveness of different ML models in clinical care, specifically in the context of chronic diseases like diabetes, cardiovascular diseases, and chronic kidney disease were examined. Moreover it analyze the challenges, of lack of standardized evaluation metrics and data governance in healthcare applications and highlight the potential of ML to revolutionize healthcare through improved diagnostic accuracy, personalized treatment plans, and better disease management strategies.

Keywords—*Chronic Diseases, diabetes, predictive analysis, machine learning.*

Introduction

A massive amount of data, including clinical information from physician notes, computerized orders, and imaging tools, have been created in the healthcare business as a result of the switch from print to computerized medical records for patients. Unfortunately, the healthcare industry has been sluggish to build technology to capitalize on such fortune because of the complexities and dispersion of these datasets, which has left a shortage of useful knowledge for better patient outcomes and care quality. This contrasts with other industries where information systems have evolved for enhanced efficiency and better returns on investments [1]. Big data and data mining applied to the usage of electronic health records in healthcare could enhance the quality of service provided. Nonetheless, there is a dearth of evaluation and synthesis of data regarding the ways in which big data has improved results [2]. Healthcare prediction has become crucial in saving lives, with the swift creation of sophisticated systems that analyze intricate data linkages. The healthcare sector is changing due to AI, which has increased the importance of machine learning and deep learning systems in diagnosing and predicting diseases. These systems simulate human perception and can diagnose difficult-to-detect diseases. Predictive analytics are essential in healthcare, as they can significantly affect disease accuracy, potentially saving lives. Reliable and effective techniques are critical for medical-purpose predictive evaluation, as are forecasts that are both on-time and accurate [3].

Dietary patterns, which include a combination of foods from various groups, have emerged as a complementary approach to understanding the relationship between diet and chronic diseases. The connection between dietary patterns—Mediterranean diet based methods to prevent hypertension (dash), Prudent, seventh-day Adventist, and Western diets—together with their affiliation with obesity, Type-2 diabetes, and cardiovascular disorder risk, asthma, and neurodegenerative illnesses are covered in [4].

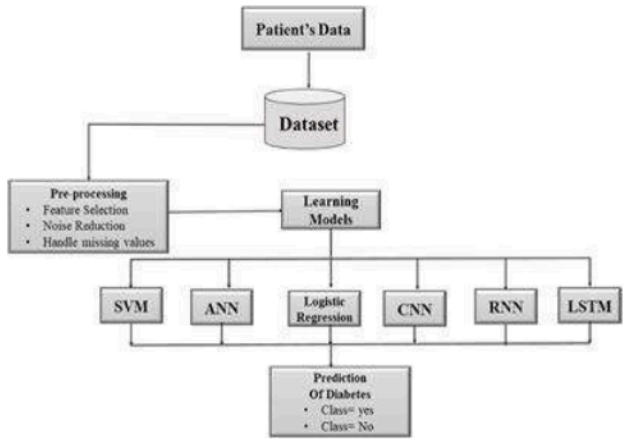

Fig. 1 Prediction of Diabetes based on machine learning approaches [3]

According to research, a diet heavy in processed and red meat, refined carbohydrates, and added sugars promotes the threat of persistent/chronic diseases, whereas plant-based totally dietary styles rich in fruits, veggies, and complete grains are useful in stopping as well as preventing them. With almost 50% of US citizens suffering from one or more chronic illness or long-term illness poses an extensive challenge to fitness care programs in the US. Pharmacist-led initiatives for controlling chronic illnesses are particularly interesting as 80 percent of medical treatments include the use of medications. Additionally, community pharmacists are among the most easily available medical professionals to provide care for a variety of populations suffering from chronic conditions. There is still confusion over the description of the care given by community pharmacists because interventions take different forms and have different outcomes in the end [5]. The relationship between dieting as a part of life and fertility is a vital cognizance for public fitness and nutrition initiatives, because it intersects with risk factors that are inclusive of obesity, depression, anxiety, and stress. Diet occupies a pivotal role inside the myriad environmental components that affect fertilization and human well-being. Research on the impact of particular nutrients, such as vitamin D, anti-oxidants, omega-3 fatty acids with long chain and dairy products on human fertility hasn't always come to a consensus. Current research on male counterpart supplements of folic acid and zinc does not demonstrate advantages in the quality of semen or rates of live births in partners gaining infertility therapy [6]. Individuals are not genetically alike, and certain genetic variants can affect metabolism, which could account for this disparity. Therefore, personalized nutrition plans based on individual genetic makeup are necessary.

II MACHINE LEARNING MODELS FOR PREDICTING CHRONIC DISEASES

AI applications in medicine are rapidly growing, with AI projects attracting more global economic speculation in 2016. AI focuses on automated diagnosis processes and patient treatment, with increased AI usage in prescriptions allowing for more time for medical experts to perform tasks that cannot be automated. By identifying intricate models and extracting medical knowledge, machine learning (ML) introduces practitioners and specialists to new concepts. Machine Learning Predictive Models can enhance decision-making in individual patient care and autonomously diagnose different diseases under clinical regulations. Incorporating ML models in medication prescription can help physicians identify pathologies and provide new medical options [7]. ML models can lower patient rate fluctuations, decrease expenditures associated with healthcare, and enhance the accuracy of medical data. Early identification and efficient care are crucial for reducing death rates from chronic diseases. Advancements in medical care have expanded electronic data accessibility, opening new doors for decision support and productivity improvements. The implementation of algorithms for machine learning (ML) along with additional Artificial Intelligence (AI) approaches to geriatric services clinical treatment for persistent medical conditions is the subject of a comprehensive literature evaluation that was carried out in [8]. The review identified 35 eligible studies, categorized into psychological disorders, eye diseases, and others. It found a absence of common metrics for ML evaluation and the requirement for data governance unique to applications in the healthcare industry. It will take further research and ML standards to verify ML's promise to enhance geriatric clinical care.

In a cohort study relying on the population, comprising 6,766 Asian individuals, the study published in [9] compared the effectiveness of logistic regression (LR) and machine learning (ML) approaches to forecast the risk of hypertension, type 2 diabetes, kidney disease that is chronic, and cardiovascular disease. Persons having CVD that is 4.0%, , CKD that is 7.0%, diabetes approximately 9.2%, and hypertension (HTN) is about 34.6% of the patient population at the conclusion of six years. For both CKD and DM, logistic regression had the largest region under the recipient's operating characteristic curve. Support Vector Machines (SVM) and Neural Networks were the most successful techniques for CVD and HTN. Among the top models were neural networks, gradient boosting machines, and logistic regression. Another study conducted in [10] presents A Random Forest (RF) regression model using actual EMR data to forecast future eGFR values. After three years of little variance, the model's average R2 of 0.95 was obtained by a the Great Plains health system. Age, BMI, and recent records are important factors that influence future variations in eGFR.Although the model only consisted of patients with numerical eGFR recordings, history eGFRs, demographic information and pertinent disease diagnoses, it has the potential to enhance clinical decision-making. One of the study in [11] seeks to avoid CKD (chronic kidney disease) by early diagnosis of the illness with machine learning techniques.The study uses a dataset of 400 patients who have 24 characteristics linked to a diagnosis of chronic kidney disease (CKD). Category techniques, which includes Support Vector machine (SVM) and Artificial Neural Networks (ANNs), are implemented. Through experimentation and parameter tuning, most beneficial parameters for SVM and ANN are determined. The final models, constructed using the best features and parameters, yield accurate results of 99.75% and 97.75% respectively. Empirical findings imply that ANN outperforms SVM, suggesting promising results. Diabetes mellitus, a chronic condition stemming from insufficient or ineffective insulin production, can result in severe complications and probably deadly consequences. Algorithms which includes SVM and ANNs are utilized for ailment prediction. The aim of the study performed by using [12] is to evaluate the overall performance of various algorithms in classifying data associated with diabetes mellitus. Results indicate that the Support Vector Machine (SVM) technique achieves a normalization accuracy of 83.54%, whereas the Artificial Neural Network (ANN) technique achieves an accuracy of 85.20%. This studies contributes to supplying specific insights into the prevalence of Diabetes Mellitus and other related health problems. Moreover, leveraging the findings from [13], healthcare experts can successfully diagnose patients with kidney illnesses and verify whether they have advanced to the chronic phase. The system employs classification modeling and machine development encompassing data guidance, collection, grouping, classification, and rule extraction. Based on multiple variables, the method has a 98.34% accuracy rate in identifying chronic renal disease. Predicting chronic kidney diseases (CKDs) is a serious task because living in large cities increases the likelihood of developing these diseases. The objective of [14] is to use MapReduce, a framework for writing predictive algorithms with map and reduce functions, to construct predictive analytics for CKD prediction. A framework for managing huge dataset samples is introduced: the iterative weighted map reduction framework. Using ensemble nonlinear support vector machines and random forests, an ensemble of deep support vector machines is created by combining various descriptors to solve a binary classification problem. Interpretability and prediction accuracy are used to assess performance. High accuracy of 93.59%, high F-measure value of 90.23%, high specificity of 89.67%, and high recall of 92.15% were attained by the suggested ESVM approach. Healthcare has undergone a transformation thanks to cutting-edge techniques like data sciences and machine learning, which have made early disease identification and better patient care possible. However, the accuracy of disease prediction is limited by incomplete medical data and region-based diseases. To address this, a disease prediction system is proposed in [15] that, by using symptoms to forecast potential diseases, can speed up diagnosis and save time. With results reaching up to 87% accuracy, the system makes reliable predictions about potential ailments using machine learning algorithms.

Table 1 Machine Learning Models for Clinical Care Prediction and Treatment of Chronic Conditions

Study Ref.	**Focus Area**	**Methods Used**	**Main Findings**	**Accuracy/Effectiveness**

[7]	Drug prescription & pathology identification	Machine Learning models	ML models store medical doctors' time and open new possibilities in pathology identification	Not specified
[8]	AI in geriatric care for persistent illnesses	thorough assessment of the literature	35 studies reviewed; need for standardized ML metrics and data governance	Not specified
[9]	Prediction of chronic diseases	ML algorithms vs. logistic regression	For CKD and DM, logistic regression worked best; neural networks and SVM were the best algorithms for CVD and HTN.	Highest AUC for CKD and DM with logistic regression
[10]	Predicting future eGFR values	Random Forest regression model	Attained a three-year average R^2 of 0.95.	Average R^2 of 0.95
[11]	Early diagnosis of CKD	ANN and SVM	ANN performed better with 99.75% accuracy	ANN: 99.75%, SVM: 97.75%
[12]	Classification of diabetes mellitus data	ANN and SVM	SVM had higher normalization accuracy	SVM: 83.54%, ANN: 85.20%
[13]	Diagnosis of chronic kidney disease	Classification modelling	98.34% accuracy rate in identifying CKD	98.34%
[14]	Predictive analytics for CKD prediction	ESVM approach using MapReduce	High performance in accuracy and interpretability	93.59% accuracy

[15]	Disease prediction system	Machine learning algorithms	System predicts possible diseases based on symptoms with up to 87% accuracy	Up to 87% accuracy

III IMPROVING DIABETES CARE THROUGH PREDICTIVE ANALYTICS

Diabetes technology offers tools for Type 1 diabetic patients to lower their risk, better manage the levels of their blood sugar, and enhance patient outcomes of acute and long-term complications. However, many patients still experience unmet goals and poor psychosocial outcomes. Big data analytics, digital health, and medical devices can all be used to improve clinical treatment, lower complications, and increase quality of life [16]. A randomized trial assessed in [17] a Stepped-Care method's impact (anticipatory analytics + personalized nurse-driven treatments) on healthcare utilization among 370 older persons utilizing a Personalized Emergency Response System. The Stepped-Care group was given an intervention for 180 days, while Controls received standard care. The Intervention group demonstrated significant reductions in a number of indicators, including entire 90-day on the number of readmission (68%), patients having Emergency Medical Services interactions (49%), 90-day on the number of readmission (76%), and total 180-day readmissions (53%), even though the primary outcome—a drop in emergency encounters—did not approach significance. Critical data has been produced as a result of developments in public healthcare and biotechnology for early detection and prevention of deadly diseases. [18] suggests a method for diabetic classification, early detection, and prediction that is based on machine learning. It uses three classification techniques: random forests, multilayered perceptron, and logistic regression, and it proposes an Internet of Things-based diabetes monitoring system. The method achieves 86.08% accuracy, outperforming other classifiers and 87.26% prediction accuracy. In [19] this pragmatic randomized trial involving 6000 participants with type 2 diabetes on basal insulin, different intensities of insulin-adherence interventions were compared. Patients were randomly assigned to receive low, moderate, or high-intensity interventions based on their predicted risk of insulin non-adherence. Contrary to the expectations, comparing to the low-intensity intervention, the heavily aimed high-level interventions failed to improve insulin persistence (with respect risk: 0.91, 95% CI: 0.77-1.06). The average glycemic control barely changed, but (unalterable level of HbA1c differences: -0.25%, 95% CI: -0.43% to -0.06%). A minor increase in hospitalizations was linked to the only partially targeted moderate-intensity intervention, although neither insulin persistence nor glycemic control were impacted (odds ratio for hospitalizations: 1.22, 95% CI: 1.06-1.41). The study conducted in [20] explores the use of six machine learning algorithms for predictive analytics in healthcare, including SVM and KNN, to predict diabetes using a PIMA Indian dataset of 768 records. The results show that SVM and KNN provide the highest accuracy at 77%, making them suitable for predicting diabetes. Nevertheless, the dataset's size and incomplete attribute values are drawbacks. In order to increase prediction accuracy, future work should concentrate on including additional techniques into the model and evaluating these models on sizable datasets with few or no absent attribute values.

IV. CONCLUSION

The advancements in artificial intelligence (AI) and machine-learning (ML) breakthroughs are an evolutionary move towards more personalised, predicting, and efficient health services in the field of medicine. Our review underscores the significant advancements and potential of these technologies in diagnosing, predicting, and managing chronic illnesses, which enhance patient outcomes and lower medical expenses. However, challenges such as data fragmentation, lack of standardization, and ethical concerns regarding data privacy and model interpretability must be addressed to fully harness the benefits of ML in healthcare. Further research and collaboration between technologists, clinicians, and policymakers are essential to develop standardized frameworks, ensure data security, and validate the effectiveness of AI and ML models in clinical settings. Ultimately, the continued evolution of these technologies holds promise for transforming healthcare systems, enhancing patient care, and advancing public health outcomes.

REFERENCES

[1] Alharthi, H.. "Healthcare predictive analytics: An overview with a focus on Saudi Arabia". Journal of infection and public health, vol. 11, no. 6, pp. 749-756, 2018.

[2] Malik, M. M., Abdallah, S., &Ala'raj, M. "Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review". Annals of Operations Research, vol. 270, pp. 287-312, 2018.
[3] Badawy, M., Ramadan, N., &Hefny, H. A. "Healthcare predictive analytics using machine learning and deep learning techniques: a survey". Journal of Electrical Systems and Information Technology, vol. 10, no. 1, pp. 40, 2023.
[4] Medina-Remón, A., Kirwan, R., Lamuela-Raventos, R. M., &Estruch, R. "Dietary patterns and the risk of obesity, type 2 diabetes mellitus, cardiovascular diseases, asthma, and neurodegenerative diseases". Critical reviews in food science and nutrition, vol. 58, no. 2, pp. 262-296, 2018.
[5] Newman, T. V., San-Juan-Rodriguez, A., Parekh, N., Swart, E. C., Klein-Fedyshin, M., Shrank, W. H., & Hernandez, I. "Impact of community pharmacist-led interventions in chronic disease management on clinical, utilization, and economic outcomes: an umbrella review". Research in Social and Administrative Pharmacy, vol. 16, no. 9, pp. 1155-1165, 2020.
[6] Franzago, M., Santurbano, D., Vitacolonna, E., &Stuppia, L. "Genes and diet in the prevention of chronic diseases in future generations". International journal of molecular sciences, vol. 21, no. 7, pp. 2633, 2020.
[7] Battineni, G., Sagaro, G. G., Chinatalapudi, N., &Amenta, F. "Applications of machine learning predictive models in the chronic disease diagnosis". Journal of personalized medicine, vol. 10, no. 2, pp. 21, 2020.
[8] Choudhury, A., Renjilian, E., &Asan, O. "Use of machine learning in geriatric clinical care for chronic diseases: a systematic literature review". JAMIA open, vol. 3, no. 3, pp. 459-471, 2020.
[9] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. "Logistic regression was as good as machine learning for predicting major chronic diseases". Journal of clinical epidemiology, vol. 122, pp. 56-69, 2020.
[10] Zhao, J., Gu, S., &McDermaid, A. "Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression". Mathematical biosciences, vol. 310, pp. 24-30, 2020.
[11] Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., ... &Olatunji, S. O. "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study". Computers in biology and medicine, vol. 109, pp. 101-111, 2019.
[12] Anggoro, D. A., &Novitaningrum, D. "Comparison of accuracy level of support vector machine (SVM) and artificial neural network (ANN) algorithms in predicting diabetes mellitus disease". ICIC Express Letters, vol. 15, no. 1, pp. 9-18, 2019.
[13] Ahmad, M., Tundjungsari, V., Widianti, D., Amalia, P., &Rachmawati, U. A. "Diagnostic decision support system of chronic kidney disease using support vector machine". In 2017 second international conference on informatics and computing (ICIC) pp. 1-4, 2019. IEEE.
[14] Prakash, S., Raja, P. V., Baseera, A., Hussain, D. M., Balaji, V. R., &Venkatachalam, K. "Ensemble Nonlinear Support Vector Machine Approach for Predicting Chronic Kidney Diseases". Computer Systems Science & Engineering, vol. 42, no. 3, 2022.
[15] Farooqui, M. E., & Ahmad, D. J. "Disease prediction system using support vector machine and multilinear regression". International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN, pp. 2347-5552, 2019.
[16] Prahalad, P., Tanenbaum, M., Hood, K., &Maahs, D. M. "Diabetes technology: improving care, improving patient-reported outcomes and preventing complications in young people with Type 1 diabetes". Diabetic Medicine, vol. 35, no. 4, pp. 419-429, 2019.
[17] Golas, S. B., Nikolova-Simons, M., Palacholla, R., op den Buijs, J., Garberg, G., Orenstein, A., &Kvedar, J. "Predictive analytics and tailored interventions improve clinical outcomes in older adults: a randomized controlled trial". npj Digital Medicine, vol. 4, no. 1, pp. 97, 2019.
[18] Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., &Sherazi, H. H. R. "Machine learning based diabetes classification and prediction for healthcare applications". Journal of healthcare engineering, 2021.
[19] Lauffenburger, J. C., Lewey, J., Jan, S., Makanji, S., Ferro, C. A., Krumme, A. A., ...& Choudhry, N. K. "Effectiveness of targeted insulin-adherence interventions for glycemic control using predictive analytics among patients with type 2 diabetes: a randomized clinical trial". JAMA network open, vol. 2, no. 3, pp. e190657-e190657, 2019.
[20] Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. "Prediction of diabetes using machine learning algorithms in healthcare". In 2018 24th international conference on automation and computing (ICAC) pp. 1-6, 2018. IEEE.