Chronic Disease Indicator Using Advanced Machine Learning Techniques

BENIL JENIFFER S,
Department of AI,
Kongu Engineering College,
Erode, India.
jeniffer.ai@kongu.edu

MADHIUKSHA S,
Department of AI,
Kongu Engineering College,
Erode, India.
madhiukshas.22aim@kongu.edu

AKSHARA R K,
Department of AI,
Kongu Engineering College,
Erode, India.
aksharark.22aim@kongu.edu

MADHUMITHA S,
Department of AI,
Kongu Engineering College,
Erode, India
madhumithas.22aim@kongu.edu

Abstract*— This study investigates modern applications of machine learning for the early detection and predictive analysis of chronic diseases, a critical medical condition. In this study, chronic diseases are analyzed using machine learning techniques. We use datasets on heart disease, cancer, and diabetes to build reliable predictive models for these chronic diseases. We use physiological signals and advanced algorithms such as Decision Trees, Naive Bayes, Random Forest, Logistic Regression, AdaBoost, SVM, and KNN to closely examine their effectiveness in detecting complex patterns indicative of impending chronic disease. The results of our research highlight the versatility of machine learning in unraveling the complexity of chronic disease prediction. These findings not only contribute to the scientific understanding of risk factors for chronic diseases but also have practical implications for timely medical interventions. By combining different algorithms and data types, our study aims to provide valuable insights that pave the way to better patient outcomes and more effective medical support strategies.

Keywords*—Machine Learning, Cancer disease, Naive Bayes, Diabetes, Random Forest, Heart Disease, Decision Tree, KNN, AdaBoost, SVM, Logistic Regression, Random Forest, Disease prediction and Accuracy.*

I. INTRODUCTION

Chronic diseases pose a significant global health challenge, contributing to a substantial burden on healthcare systems and impacting the quality of life for millions of individuals. These diseases, such as cardiovascular conditions, diabetes, and certain types of cancers, often require long-term management and care. Early detection and proactive intervention play a crucial role in mitigating the progression of chronic diseases, improving patient outcomes, and reducing healthcare costs. Machine learning, a subset of artificial intelligence, has demonstrated remarkable efficacy in various healthcare applications. This includes the application of Naive Bayes and Random Forest algorithms, which exhibit the potential to analyze the risks. This report aims to provide a comprehensive overview of the application of machine learning in chronic disease indicators. It will delve into the various ways in which machine learning models are employed for early detection, predictive analytics, personalized treatment plans, and population health management. In the subsequent sections, we will delve into the related works that lay the foundation for our research, present the detailed methodology employed for integrating machine learning techniques, and provide comprehensive insights into the proposed work's potential implications for healthcare interventions. With this research, we aim to pave the way for a more proactive and effective approach to tackling the challenges associated with chronic disease, ultimately improving patient outcomes and advancing medical diagnostics.

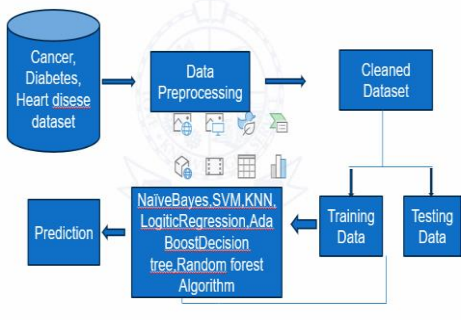
II. RISK FACTORS

Chronic diseases, known as non-communicable diseases, usually appear in middle age, less often in teenage years, after long exposure to an unhealthy lifestyle associated with tobacco use, lack of regular physical activity, and the consumption of foods high in saturated fats, sugar, and salt, typically "fast food". This lifestyle leads to an increased number of risk factors such as hypertension, dyslipidemia, diabetes, and obesity, which act independently and synergistically. These risk factors are often undiagnosed or inadequately treated in health services that focus on treating acute conditions. The current burden of chronic disease reflects the cumulative effects of unhealthy lifestyles and the resulting risk factors across the lifespan. Some of these influences are already present before a child is born.

III.METHODOLOGY

The chronic disease prediction methodology used in this study includes an algorithmic implementation. The diverse dataset, which includes physiological and textual data, was carefully curated and pre-processed to ensure the quality and consistency of the data. In addition, the results of the seven algorithms are not only quantitatively compared using metrics, but also qualitatively analyzed to decipher the nuances in their predictions. The methodology thus goes beyond a purely technical process and promotes a deeper understanding of the interplay between algorithms and data modalities. It provides valuable insights for the further development of predictive models in the field of chronic disease prediction and healthcare applications.

IV.PROPOSED WORK

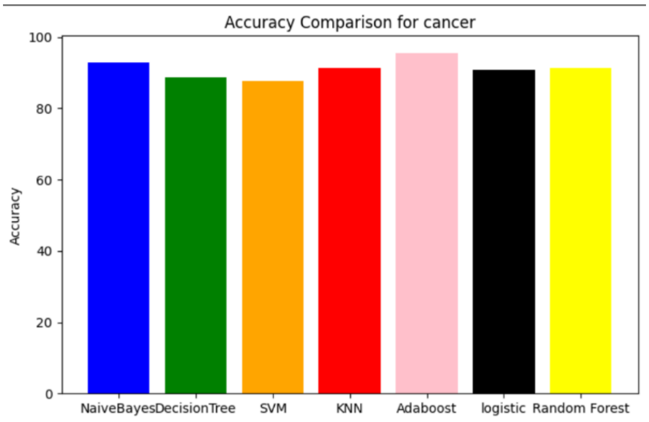


A predictive model for early diagnosis of chronic diseases involves strengthening machine learning algorithms on relevant health data. The proposed system integrates patient records, lifestyle factors, and genetic information to predict the likelihood of developing a chronic disease. It uses predictive analytics to identify patterns and risk factors that enable vigorous intervention and personalized preventive measures. The system aims to improve early detection to ultimately improve patient outcomes and reduce healthcare costs through targeted timely interventions. The study places a strong emphasis on quantitative and qualitative analysis of the algorithms, offering nuanced insights into their strengths and limitations. Ultimately, the goal is to gain actionable insights for real-world healthcare applications through a better understanding of predictive modeling of chronic diseases.

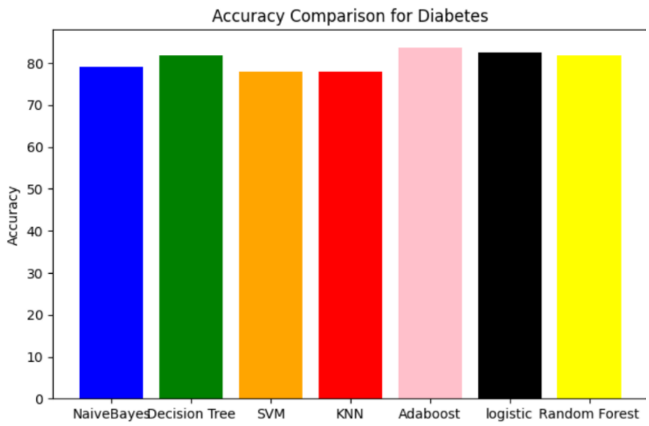
V. RESULT

Disease	Algorithm	Accuracy
Cancer	Naive Bayes	92.98
Cancer	Decision Tree	88.6
Cancer	Random Forest	91.23
Cancer	SVM	87.72
Cancer	KNN	91.23
Cancer	AdaBoost	95.61
Cancer	Logistic Regression	90.74
Diabetes	Naive Bayes	79.22
Diabetes	Decision Tree	81.82
Diabetes	Random Forest	81.82
Diabetes	SVM	77.92
Diabetes	KNN	77.92
Diabetes	AdaBoost	83.77
Diabetes	Logistic Regression	82.46
Heart Disease	Naive Bayes	74.07
Heart Disease	Decision Tree	85.19
Heart Disease	Random Forest	85.19
Heart Disease	SVM	68.52
Heart Disease	KNN	74.07
Heart Disease	AdaBoost	85.19
Heart Disease	Logistic Regression	83.33

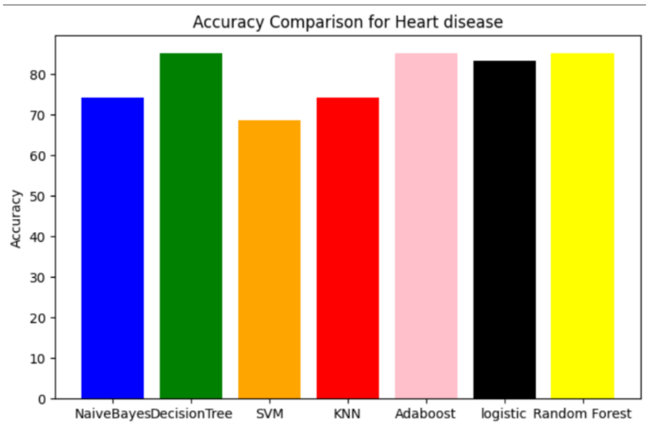
Bar plot for Cancer



Bar plot for Diabetes



Bar plot for the Heart Disease



The predictive performance of different machine learning algorithms for chronic disease prediction is evaluated using key metrics. For the cancer dataset, Naive Bayes achieves an accuracy of 92.98%, Random Forest follows closely behind with 92.11%, Decision Tree shows a good accuracy of 88.6%, SVM with an accuracy of 87.72%, KNN with an accuracy of 91.23, AdaBoost with an accuracy of 95.61, Logistic Regression with an accuracy of 90.74%. For the diabetes dataset, Random Forest achieves a high accuracy of 83.77%, Naive Bayes an accuracy of 79.22% Decision Tree a good accuracy of 81.82%, SVM an accuracy of 77.92%, KNN an accuracy of 77.92, AdaBoost an accuracy of 83.77, Logistic Regression with an accuracy of 82.46%. For the heart disease dataset, Random Forest and Decision Tree achieve an equal and higher accuracy of 85.19%, Naive Bayes an accuracy of 74.07%, SVM an accuracy of 68.52%, KNN an accuracy of 74.07%, AdaBoost an accuracy of 85.19%, Logistic Regression an accuracy of 83.33%. These results provide a quantitative overview of the effectiveness of each algorithm in the context of chronic disease prediction, with Random Forest emerging as the best-performing algorithm in this evaluation.

VII.CONCLUSION

To summarise, our disquisition of machine knowledge algorithms for habitual complaint vaticination provides compelling perceptivity into their performance. The results show a competitive terrain with a advanced delicacy of95.61 for cancer for the AdaBoost algorithm,83.77 for diabetes, and85.19 for heart complaint. These results illuminate the versatility of machine knowledge in the field of habitual complaint vaticination, offering a range of algorithms with remarkable rigor. The nuanced analysis of the algorithms' results provides precious considerations for concluding applicable models in real- world healthcare operations. As the field continues to evolve, this exploration contributes to the ongoing converse on the use of machine knowledge for accurate and timely vaticination of habitual conditions, toameliorate issues for cases and healthcare interventions.

VII.REFERENCE

1)HametP., TremblayJ. Artificial intelligence in drug. Metabolism. 2017; 69S36 – S40. doi10.1016/j.metabol.2017.01.011.(PubMed)(CrossRef)(Google Scholar).
2)JohnsonK.W.,SotoJ.T.,GlicksbergB.S.,ShameerK.,MiottoR.,AliM.,DudleyJ. T. Artificial intelligenceincardiology.J. Am. Coll. Cardiol. 2018; 712668 – 2679. doi10.1016/j.jacc.2018.03.521.(PubMed)(CrossRef)(Google Scholar).
3) BiniS. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing What Do These Terms Mean and How Will They ImpactHealthCare?J. Arthroplast. 2018; 332358 – 2361. doi10.1016/j.arth.2018.02.067.(PubMed)(CrossRef)(Google Scholar).
4)KotsiantisS.B.,ZaharakisI., PintelasP. Supervised machine knowledge A review of type ways. Emerg. Artif. Intell.Appl.Comput.Eng. 2007; 1603 – 24.(Google Scholar). 5)DeoR.C. Machine knowledge in Medicine. Gyration. 2015; 1321920 – 1930. doi10.1161/CIRCULATIONAHA.115.001593.(PMC free composition)(PubMed)(CrossRef)(Google Scholar).
6)BattineniG.,SagaroG.G.,NaliniC.,AmentaF.,TayebatiS.K. relative Machine-Learning Approach A Follow- Up Study on Type 2 Diabetes prognostications byCross- Validation styles. Machines. 2019; 774. doi10.3390/ machines7040074.(CrossRef)(Google Scholar). 7)PolatH.,MehrH.D., CetinA. opinion of habitual order Disease rested on Support Vector Machine by pointSelectionstyles.J. Med. Syst. 2017; 4155. doi10.1007/ s10916-017-0703-x.(PubMed)(CrossRef)(Google Scholar).