



# A Comprehensive Survey of Machine Learning Algorithms for Multi-Disease Prognosis

Sharmila Rathod<sup>1</sup>, Aryan Panchal<sup>2</sup>, Jash Panchal<sup>3</sup>, Ashlesha Padvi<sup>4</sup>

<sup>1,2,3,4</sup> Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, India.

## How to cite this paper:

Sharmila Rathod<sup>1</sup>, Aryan Panchal<sup>2</sup>, Jash Panchal<sup>3</sup>, Ashlesha Padvi<sup>4</sup>,  
"A Comprehensive Survey of Machine Learning Algorithms for Multi-Disease Prognosis", IJIRE-V5I03-188-189.

Copyright © 2024 by author(s) and 5th Dimension Research Publication. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>

**Abstract:** In this comprehensive analysis, study of various machine learning-based systems for multi-disease prognosis has been conducted. Predictive machine learning algorithms are used extensively in the domain of medical science thereby leading to a considerable improvement in accurately predicting a disease. The timely identification and accurate measurements of these conditions hold the potential for a methodical and efficacious treatment. As research solidifies, the possibilities of computing methods to both optimize and enhance comparative systems seems vast. This comprehensive review aims at providing an intricate and highly detailed analysis of numerous machine learning algorithms and the functioning working environment specifically focused on prognosis of diseases such as myocardial infarction, diabetes and chronic kidney disease. The study thus offers an amalgamation of the most recent medical surveys, thereby contributing to the ongoing research in the field of medical science.

**Key Word:** machine learning, myocardial infarction, diabetes, chronic kidney disease, efficacious treatment, medical surveys.

## I. INTRODUCTION

Recent technological advancements in the field of health sciences and medicine have produced life-saving outcomes. From prognosis to diagnosis, software systems in general and machine learning algorithms in particular have managed to deliver accurate results. The system of taking user inputs on various parameters like BMI, Blood Sugar Level, age etc for detecting diabetes mellitus (DM) and further analysing and processing the collecting information through the Machine Learning model has been practiced in several researches conducted across the globe. Collected datasets have a comparatively high number of quality attributes out of which only a few have to be selected for research purposes. Creation of a mixed model to select the best accuracy is widely being used and is achieved using the classifier models. After the comparative analysis, various algorithms provide a varied accuracy level of the result obtained. For example, SVM algorithm may give highest accuracy for myocardial infarction and Random Forest algorithm may provide an accurate result for cancer related ailments. Selecting particular features from the dataset, modifying the algorithm and performing efficient pre-processing of data are keys which can open doors to further research as far as disease prognosis is concerned.

## II. LITERATURE SURVEY

[1] The conference paper presented by Akkem Yaganteswarudu titled "Multiple Disease Prediction Model Using Machine Learning and Flask API". In multiple pathology model, more than one disease can be predicted simultaneously. So users don't have to repeat many models to predict disease. Since many diseases are predicted simultaneously, it will shorten the duration and reduce mortality rate. Use a variety of diagnostic tools using machine learning algorithms, tensor streaming, and Flask API. Python pickling is used to save the model, while python unpickling is used to load saved data when necessary. Some existing systems exploit various flaws while identifying viruses. Therefore, consequences of this disease will not be recognized. When creating a new virus, developers need to prepare Python options that will protect the code. While using this Flask API, developers can load a database to store samples. When the user wants to check the patient's health, a specific disease can be predicted or if the report has indicators to predict various other diseases, the Scan will focus on that disease.

[2] As stated in the paper titled, "Multiple Disease Prediction using Machine Learning and Streamlit", Each prediction is first compared and out of that the most accurate prediction is displayed as the final output. The working of it is as follows, there are multiple datasets for each disease and the databases which are acquired. The pre-processing methods that are used in this are label encoding and by creating a mixed model to select the best accuracy using the classifier module. Each model is first trained against its training set then accuracy is calculated and compared to each other and once it is completed, whichever is the best one is selected.

[3] The research paper "Disease Detection Using Machine Learning Algorithms" that is published by Sneha Grampurohit and Chetan Sagarnal aims to develop and use various knowledge mining methods in various global applications such as business, medicine and biosciences, thereby contributing to the use of knowledge mining methods in various global applications such

as business, medicine and biosciences, public health, biomedical, etc. The use of data mining techniques that use such techniques in the machine learning environment to extract useful information from datasets across domains. Accurate analysis of medical records facilitates early prediction of disease, patient care as well as community service. Many applications, including disease prediction, have effectively exploited current machine learning techniques. By assisting physicians in early disease prediction and diagnosis, classification systems developed utilizing machine learning algorithms are intended to contribute to the solution of health-related issues.

[4] The primary utilization of the categorization approach is to initially prognosticate the target category precisely for every instance existing in the dataset. The principal and most significant emphasis of this document is to prognosticate Chronic-Kidney-Disease and its application for categorization in the realm of medical bioinformatics. It categorizes the dataset and subsequently ascertains which algorithm is superior for diagnosis and prognosis of Chronic-Kidney-Disease. The primary aim of the research is to scrutinize the data from a persistent-kidney-disease (CKD) dataset employing the categorization methodologies to precisely prophesy the category in each instance.

[5] In the article entitled "Enhancing Classification Performance through Pre-Processing and Machine Learning Algorithms on NSL-KDD Dataset" authored by Datta H. Deshmukh and Tushar Ghorpade, the authors delve into the notion of taxonomy, which essentially refers to a comprehensive category encompassing an extensive list of classes. The primary focus of this scholarly paper is to elucidate a meticulously designed step-by-step procedure that leverages pre-processed data, encompassing pivotal elements such as feature selection and decision-making techniques. It is important to highlight that when dealing with voluminous files, the sole information employed in this process is a customized selection and subsequent size reduction. In the initial stages, a myriad of theories are meticulously selected through the utilization of selection algorithms, and these specific endeavours significantly contribute to the overall enhancement of the accuracy of the classification process. Subsequent to this crucial step, an array of tools are employed, including Naive Bayes, Hidden Naive Bayes, and NBTree. A noteworthy advantage of the latent negative Bayes technique lies in its ability to relax the independence assumption that is inherently present in the naive Bayes method, thereby resulting in a marked improvement in the overall effectiveness of the data mining model. Additionally, the subsequent technique implemented is NBTree, which seamlessly combines the decision tree classifier with the naive Bayes classifier, culminating in an enhanced accuracy of classification and a notable reduction in the error rate of the isolate. Following the completion of the aforementioned steps, it becomes imperative to meticulously evaluate the output of the proposed process, discerning whether it is positive, negative, negative, or false. Based on the outcomes of this meticulous assessment, the accuracy and error of each classification are subsequently calculated and comprehensively evaluated.

### III.CONCLUSIONS

In this paper, we have comprehensively deliberated upon the distinct methodologies that are employed for the prognosis of multiple diseases. Commencing with the present survey, it is of utmost significance to thoroughly investigate the intricate workings of algorithms. The comparison of diverse algorithms has demonstrated promising potential in yielding commendable outcomes. This expeditious examination can prove to be exceedingly advantageous for both researchers and healthcare professionals alike.

### IV.ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal and Prof. Sunil P Khachne, H.O.D of Department Computer Engineering of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our project work. This project bears on imprint of many people. We sincerely thank our project guide Dr. Sharmila Gaikwad for her guidance and encouragement in carrying out this synopsis work. Finally, we would like to thank our colleagues and friends who helped us in completing project work successfully.

### References

- [1] Akkem Yaganteeswarudu, *Multi Disease Prediction Model by using Machine Learning and Flask API*, Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record.
- [2] Laxmi deepthi Gopisetti, *Multiple Disease Prediction System using Machine Learning and Streamlit*, IEEE, 2023.
- [3] Sneha Grampurohit, Chetan Sagarnal "Disease Prediction using Machine Learning Algorithms" IEEE International Conference for Emerging Technology (INCET), Jun. 2020.
- [4] Lambodar Jena,Ramakrushna Swain "Work-in-Progress: Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers" International Conference on Information Technology. 2017
- [5] Datta H.Deshmukh, Tushar Ghorpade, "Improving Classification Using Preprocessing and Machine Learning Algorithms on NSL-KDD Dataset", IEEE, 2015.
- [6] "Integrated Xception- Random Forest Model for the Detection of Rheumatoid Arthritis in Hand Thermograms", 12th IEEE International Conference on Communication Systems and Network Technologies.
- [7] Rojaramani D, 'Tweak Myocardial Infarction (MI) prognosis method using RFGS ( Random Forest Grid Search ) optimisation for Machine Learning and Hyper Parameters.' in THE INDIAN JOURNAL OF TECHNICAL EDUCATION VOL. 45 JAN-MARCH, 2022
- [8] Kedar Pingale, Sushant Surwase, Vaibhav Kulkarni, Saurabh Sarage, Prof. Abhijeet Karve, "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 12 | Dec 2019.
- [9] Imesh Udara Ekanayake, Damayanthi Herath, "Chronic Kidney Disease Prediction Using Machine Learning Methods", IEEE, 2020.