Intelligent Healthcare: AI And Machine Learning Approach in Chronic Disease Detection

Jignasa Patel[1], Dr. Kruti Dangarwala[2], Jalpa Patel[3]
[1,2,3]*Computer Science & Engineering Department SVMIT Bharuch, India*

Abstract— **The prediction of human diseases constitutes an essential component of human existence. The process of disease prediction entails estimating the likelihood of patients' illness by analyzing the interplay the symptoms of patient. Vigilantly patient health monitoring status and pertinent data during the initialization of assessment can significantly aid medical practitioners in effectively addressing the patient's health concerns. Recent progressions in Machine Learning (ML) approaches have resulted in considerable improvement in the identification and prediction of health crises, demography of disease, disease progression, and immunological responses among others. The objective of this article is to represent a comprehensive overview of various algorithms like support vector machine, Naïve Bayes, K- nearest Neighbor, Random Forest, Logistic Regression and decision tree of machine learning that help in the detection of various diseases. This study systematically reviews a multitude of prior research endeavours that have used various machine learning algorithms for the identification of diverse diseases within the domain of healthcare. It is particularly critical to diagnose individuals with chronic diseases at the early possible stage. This review has concentrated on evaluating the application of ML techniques for predicting chronic diseases, such as heart disease, cancer, and liver ailments. In this review, numerous previous studies focusing on chronic disease prediction were examined, and we assessed and compared the accuracy of various ML algorithms in detecting breast cancer.**

Index Terms— **Artificial intelligence, Machine Learning, Healthcare, K-Nearest Neighbor, Random Forest, Naïve Bayes, Logistic Regression, Decision Tree, Chronic Diseases Prediction.**

I. INTRODUCTION

Disease prediction is a vital part of human life. The healthcare and medical sector are more demanding sector of work today. When data mining methods are used properly, we can extract valuable information from large database and that helps doctor to take early decision and improve patient health services [1]. Disease like heart disease, diabetes, liver disease, cancer, malaria cause significant effect on individual person's health and sometimes might lead to death if it is ignored [1]. Chronical disease is a last long and take long time to heal it cannot be cured but manage with daily good treatments [2]. Today's economic growth and urbanization have resulted in major life style changes that affect the people's health [2]. Business procedure has change significantly due to artificial intelligence and machine learning that affect people's daily lives and similar changes are expected in healthcare and medicine industry [3]. The main function of machine learning in healthcare have to provide assistance and supportive role to physicians to carry out their duties, spotting trends in healthcare and create disease prediction model [3]. Disease diagnosis using machine learning techniques is time efficient and affordable that is illustrated by many practitioners and researchers [4]. Traditional process for disease diagnosis is costly, take more time and that require human involvement even it is limited by individual's ability, this kind of constraint can be overcome with help of ML based system [4]. Health care data, including pictures such as X-rays and MRIs and tabular data such as patient conditions, age, and gender are used to construct machine learning base systems [4]. While a lot of research has been conducted regarding specific Chronical Disease, where most of the Scholars have looked at several facets and consequences of that particular disease [12]. Here, in this article we mainly focus on the prediction of chronical disease including liver disease, cancer, and heart disease and also compare accuracy of different ML algorithm for breast cancer detection using machine learning algorithms.

II. OVERVIEW

The advancement utilization of various significant data mining methodologies across a multitude of real-world application domains (such as Industry, Healthcare, and Biosciences) have facilitated the implementation of these methodologies within machine learning frameworks, thereby enabling the extraction of pertinent information from specialized datasets within healthcare communities, biomedical sectors, and similar fields. The precise examination of medical databases yields advantages in the early disease detection, enhancement of patient's care, improvement in community health services [1]. Malaria, Diabetes, Migraine, Chickenpox and other health related issue have major impact on person's health [1]. Recent progressions in the domains of Artificial Intelligence (AI) and ML technologies have advancements in the prediction and identification of health issues, demographics of disease, immune responses, among others [3]. Human disease prediction constitutes an essential component of human existence. The early identification of diseases in individuals represents a critical phase in the therapeutic process. Historically, this responsibility has been predominantly managed by medical practitioners. Consequently, the healthcare sector relies heavily on innovative advancements to enhance logistical efficiency. Innovation serves as the cornerstone of the medical field. It is the catalyst for the development of novel treatments, cures, and medical interventions. [5].

A. Artificial Intelligence

AI, ML and Deep Learning (DL) refer to distinct sets of algorithms and learning procedures. Any electronic intelligence that learns from the intellect of humans is referred to as artificial intelligence [3]. In healthcare AI is not new idea it has been wide spread utilized and it has a capacity to greatly improve overall patient care treatment effectiveness, and diagnostic accuracy [21]. In medical domain its primary goal is to develop algorithm to determine disease diagnosis is correct or not as explain a person's symptoms, It is usually difficult because a lot of symptoms and indicators are unclear and same like other disease only qualified medical professionals can make a diagnosis [4]. AI is currently used in the development of diagnostic tools. For example, AI algorithm can analyze CT scans and X-rays for signs of illness or damage [21]. By analyzing huge amount of medical data, AI can help doctor to understand how genetics, environment, and lifestyle impact on patient's health [21].

B. Machine learning

ML is a subpart of Artificial Intelligence technology. ML uses mathematical and statistical approaches to analyse samples of data and find key conclusions. Fundamental idea of machine learning is to enable machines that learn without hard programming and trained from data to make judgement and project outcome based on assign task [4]. Many laborious duties may now be finished quickly and with little effort thanks to machine learning technology. Due to growth of computer power and data capacity it is easy to train ML models to make nearly unblemished predictions [4]. Numerous articles present different kinds of machine learning techniques. Today's world ML is used everywhere from mobile phones, computers, robotics and healthcare industry. Various ML techniques are used to detect chronical disease like breast cancer, heart disease, and liver disease. Early and more accurate diagnosis, individualized treatment regimens, faster medication discovery, better monitoring of patients, and hospital services are all made possible by ML techniques. To forecast disease and health outcomes, ML algorithms analyze huge amounts of data, including health records and photographic images, ultimately all it leading to improvement in patient care and boosts operational effectiveness.

III. OBJECTIVE

The primary aim of this review is to furnish an examination of various methodologies pertaining to machine learning as well as to offer insights concerning machine learning-driven disease diagnosis that will assist and empower practitioners in selecting the most suitable machine learning techniques. we summarize different machine learning algorithms used in disease diagnosis and find accuracy of chronical disease.

IV. MACHINE LEARNING ALGORITHM

The Machine learning, algorithms are classified mainly in three categories such as supervised learning, unsupervised learning and semi supervised learning.

Fig-1 shows classification of ML algorithms into several subgroups based on different learning methods. Some of the popular ML algorithms are Support Vector Machine (SVM), Logistic Regression, Random Forest, Naïve Bayes and Decision Tree [4]. A Comparative analysis of these machine learning algorithm is presented in table I.

A. Support Vector Machine (SVM)

SVM is used to classify diseases according to patient's symptoms. SVM model is effective in disease prediction but, it takes more time to predict disease [5]. Also, this method cannot improve the model's accuracy. The method's disadvantage is that it uses a hyper plane to classify objects, which is only partially successful so, only two groups of sample data can be accurately classified using the hyperplane [5][7]. Now a days, the medical field needs more than two groups (diseases) in order to identify the symptoms that correlate to each disease [5][7].

B. K- Nearest Neighbor (KNN)

The K- Nearest Neighbor is a nonparametric classification approach. It is well-known for being easy to use and efficient. K plays an important role in classifying unlabeled data [6]. The correctness of the method is mostly determined by calculating accurate distance. Determining parameter K is one of the major steps.[6] Sometimes it's unclear what kind of distance to consider and which feature will best for work. It is a lazy learning method that uses data to classify input rather than learning from training data [6]. KNN is costly and requires more storage for huge dataset. The K-Nearest Neighbors algorithm, which is sensitive to missing and noisy data, and allocation the data point to the class to which the predominant number of the data points are affiliated [5][7]. In research [5][7] they have taken into account a number of parameters for measure, including the patient's gender, age group, and symptoms, although the accuracy of machine learning models is lower, they have demonstrated good accuracy in a number of situations, including the prediction of heart risk and diabetes.

C. Naïve Bayes (NB)

NB is Statistical and Probability method-based classification algorithm. NB is standard algorithm of machine learning, because of how simple it is to enable [8]. Every feature must equally affect the ultimate choice. This simplicity is equivalent to computational efficiency, which makes the NB approach suitable for a variety of domains. Prior, posterior, and class conditional probabilities are the essential elements of the NB classification. In paper [7], they describe NB for limited diseases such as Diabetes, Malaria, Jaundice, Dengue, and Tuberculosis method not worked on a large dataset to predict large numbers of diseases. Among its numerous advantages are its ease of use and suitability for large datasets [8]. It might be applied to both binary and multiclass classification issues also it can be applied to both continuous and discrete data and requires less training data. This algorithm can be applied to document classification and spam email filtering [8].

D. Logistic Regression

Calculated Regression was predominantly applied in the realm of natural research and applications during the mid-20th century. One supervised learning method that is useful for solving binary classification issues is logistic regression [9]. It employs a mathematical model in conjunction with the logistic function to facilitate binary classification tasks. This algorithm is capable of accommodating an arbitrary number of numerical as well as categorical variables. Moreover, it introduces a discrete output variable that ranges between 0 and 1[9].

Fig.1. Different Machine Learning Algorithms [4]

E. Random Forest

Random forests also known as random decision forests. This is an ensemble learning approach for classification, regression, and other problems that work by building a large number of decision trees during the training phase and producing the mean prediction for regression or the dominant class for classification from each of the trees [9]. The likelihood of algorithm to overfit their training datasets is effectively addressed by random decision forest and there is a direct relationship between the overall quantity of trees in this group and the output that can be produced. Random forests give another degree of unpredictability to the training process in order to improve prediction accuracy [9]. Random forest algorithm driving force is that it has the capacity to work with categorical variables and continuous variables with respect to classification and regression [5].

Table I: Comparative analysis of ML Technique

ML Technique	Advantage	Disadvantage
Support Vector Machine (SVM) [5][8]	• Faster Execution • Less Space complexity • Work on recent data. • Work on both linear and nonlinear datasets. • Probability of overfitting is less.	• Require more time to predict disease. • Do not consider medical history as input data. • Classifying object using hyperplane which is partially effective. • Not Suitable for Multi-parameter. • Performance goes down with large dataset. Do not work well for noisy dataset
K-Nearest Neighbors (KNN) [6][8][25]	• It is simple, scalable and comprehensible. • Interpretation is easy and require less time for calculation. • Prediction capability is high, so it makes more effective and efficient. • It works efficiently for both classification and regression.	• For huge dataset difficult to determine K. • Classification of unknown records are costly. • For irrelevant data algorithm is Highly sensitive. • Computation is very high.
Naïve Bayes [5][8]	• Highly scalable • Easy to work for large collection of datasets • distinct and continuous data can be managed easily. • Work for both classification binary as well as multiclass. • No more sensitive for noisy features.	• computationally demanding, especially for models with a large number of variables. • Only for independent features it works • Accurately.
Logistic Regression [5][8]	• It makes assumption about • Distribution. • used for both regression • and classification	• It requires less multi-collinearity • due to Over-Fitting problem. • Solving a nonlinear problem is difficult.
Random forest [5][8][9]	• It works for both classification as well as regression • Mitigates overfitting problem of decision tree.	• It needs more time to train the dataset. • Complexity
Decision Tree [8][10]	• It can work regression and classification • Manage both numeric and categorical data • Possesses very low complexity • Beneficial for Data exploration • Less data cleaning required • No restriction on data type • Non-Parametric approach	• Overfitting is problem when tree built repeatedly • Interpretation is hard when tree is large. • Not suitable to work with continuous variables.

F. Decision Tree (DT)

The DT algorithm, classified within supervised learning algorithms, is predominantly favored for addressing challenges in classification nevertheless, it is equally applicable in both classification and regression scenarios [10]. This algorithm is made up of internal nodes that describe the branching structures, datasets that represent the decisions rendered by the algorithm, and leaf nodes that denote the outcomes [10]. two nodes exist in decision tree, decision node that takes a decision and it contain many branches, another is output of decision node known as leaf node and it does not contain any further branches [10]. The DT can handle categorial as well as numerical data, further more effectiveness of tree is not affected by nonlinear relationship among parameters [8][22]. In DT data preprocessing is not required. The probability of overfitting may occur when tree built repeatedly [8][23]. In paper [24] author perform experiment using decision tree to detect breast cancer, leaf nodes are divided into two categories: Benign or Malignant. Rules will be redefined among selected dataset to decide whether the tumor is benign or malignant based on clump thickness [8][24].

V. MACHINE LEARNING APPLICATION IN HEALTHCARE

ML algorithms play a crucial role to recognize complex patterns in large and complex datasets. Capability to find complex pattern is particularly advantageous for clinical applications, especially for individual's dependent on advanced proteomics and genomics measurements. Algorithms are frequently employed in the diagnosis of various medical conditions. Within medical contexts, machine learning algorithms are capable of generating superior decisions concerning patient treatment plans by recommending the implementation of effective healthcare systems [11]. Healthcare management is employing this methodology to predict waiting periods for patients in emergency departments seeking admission. These models incorporate variables such as patient demographics, pain levels, emergency department records, and even the spatial configuration of hospital facilities to ascertain wait times [11]. Through the utilization of prognostic models, healthcare facilities will consider admission to hospital Thus, machine learning application could beneficial to patients in terms of costs, enhancing precision, mitigating the experience of scarcity in healthcare services [11].

A. Chronic Diseases Prediction Using ML

Any illness or condition that last for minimum duration of three months and has potential to have serious long-term effect is defined as chronical disease. It is more prevalent among older adults and can generally be managed but not completely eradicated. Common manifestations of chronic disease include cancer, cardiovascular disease (CVD), diabetes, neurological disorders, hepatic diseases, stroke, and arthritis. Some of chronical disease like stroke, cancer, cardiovascular, hypertension, diabetes requires early prevention and diagnosis as per recommendation of WHO (world health organization) [13]. One preventive strategy that can be implemented is the prediction of chronic diseases through machine learning, utilizing individual medical records or general health assessment results [13].

B. Liver Disease

The liver is main largest internal organ of the human body that perform two primary functions blood filtration and digestion [26]. As food traverses the digestive system, the liver distinguishes beneficial components from detrimental ones. Additionally, it produces bile, which facilitates purges and digestion toxic substances from the body. functional capacity of liver declines when fibrous scar tissue progressively replaces healthy hepatic tissue. If not addressed promptly, LD can potentially culminate in liver failure and cancer. Various potential etiologies for LD exist, including infections, suboptimal dietary habits, substance abuse, alcohol dependency, and exposure to toxins. Furthermore, genetic predispositions for LD are also present. The development of more accurate predictive models utilizing a diverse array of machine learning methodologies is gaining traction in light of the increasing work with ML in the sector of healthcare. Disease diagnosis process, prevention or treatment could be greatly enhanced by early identification of risk factor of liver disease. As in [14] authors described various methods to diagnosed Non-Alcoholic Fatty Liver Disease (NAFLD), they take training sets of 10,373 participants from which Ultrasound tested 4,018(38.74 %) participants found with NAFLD, as well 1,860 (37.64 %) diagnosed with

NAFLD with testing set of 4,942 participants. Remarkable difference is measure in with and without NAFLD. In training and testing sets, measure remarkable difference between with and without NAFLD with respect to parameters like age, BMI, waist circumference (WC) [14]. Various ML algorithms like XG Boost, LR, SGD, AVM, CNN, MLP, LSTM used to test NAFLD and also check efficiency of methods in training set [14].

C. Heart Disease

In the world most known illness is heart disease, due to this disease many people lost their lives [16]. We can save the people lives with early detection of disease. Use Machine learning algorithm is most convenient way to detect disease. Group of diseases like heart valve problem, heart failure, coronary artery problem all are considered into heart condition or disease [16]. Ahmad Ayid Ahmad et al. [16] perform the implementation in their proposed method and compared result with algorithms of machine learning such as ANN, Decision Tree, AdaBoost, SVM, according their experiments they find the classification accuracy was 98.08%, 97.43%, 97.84%, 98.09% respectively that shows with 98.09% accuracy SVM classifier model perform more accurate result [16].

D. Uses of ML in Cancer Prediction

Cancer constitutes a preeminent factor contributing to both mortality and morbidity on a worldwide. In spite of advancements in the domains of diagnosis, prognosis, and therapeutic interventions for cancer patients, the quest for personalized and data-driven healthcare continues to present formidable challenges [17]. The incorporation of artificial intelligence, which is utilized for the prediction and automation of various cancer types, has emerged as a viable approach for augmenting the precision in healthcare services and optimizing patient result in diagnosis. AI application within oncology field encompasses risk assessment, early detection, prognostic evaluation, and treatment selection based on extensive knowledge repositories. Special subset of AI is machine learning which allow computers to acquire knowledge through training datasets, has demonstrated particular efficacy in predicting a range of cancer types such as brain, breast, lung, liver, prostate cancers. Machine Learning algorithms have shown superior capability to find accuracy in prediction of cancer relative to medical practitioner. Additionally, these emerging methods possess the potential to enhance diagnostic precision, prognostic capabilities, and patient's perceived health suffering from a variety of medical conditions, extending beyond the realm of oncology [17].

VI. TYPES OF CANCER DATA ANALYSIS

When undertaking analytical procedures, diverse categories of data are accessible within the domain of oncology, which are frequently employed to derive insights and facilitate informed decision-making. The nature of the data employed is contingent upon the precise research objectives and the resources at hand. Presented below are several various types of data used for analysis of cancer as referenced in this survey:

1. Medical Information/Clinical Data

This encompasses data pertinent to patients, including demographic variables, historical medical context, symptomatology, treatment documentation, laboratory findings, pathology assessments, and Clinical result. Medical information or clinical furnishes critical understanding of patient's data attributes and the trajectory of disease advancement [18].

2. Genomics and Genetic Engineering

In the field of genetic engineering adaptive DNA system has developed which is known as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)[3]. The ML application in genetic engineering has major inputs to fight against COVID-19[3].Genomic data pertains to the analysis of the genetic makeup of neoplastic cells, incorporating DNA sequencing information, gene expression profiles, and genetic polymorphisms, This information is instrumental in the identification of genetic aberrations, gene expression trends, and prospective biomarkers for the purposes of cancer diagnosis, prognosis, and the selection of therapeutic interventions [18].

3. Imaging Data

Imaging data is generated through various source like 'CT (computed tomography)', 'MRI (magnetic resonance imaging)', 'PET (Position emission tomography)', X-ray medical imaging techniques

[18][19]. These imaging outputs deliver comprehensive information regarding the tumor's location, dimensions, morphology, and characteristics, thereby facilitating diagnosis of cancer, staging, and planning of treatment [18][19].

4. Omics Data

Omics data includes extensive molecular datasets, including transcriptomics, proteomics, metabolomics, and epigenomics [18]. This information offers understanding of the molecular alterations taking place in cancer cells and may assist in pinpointing new therapeutic targets and biomarkers [18].

5. Electronic Health Record (EHR)

Electronic Health Records comprise a comprehensive information of patient including medical data, patient's medical history, treatments and outcomes [18]. EHR data mining facilitates the examination of extensive patient cohorts and identify the trends and patterns related to cancer incidence, treatment efficacy, and outcomes of patient [18].

6. Dataset Available Publicly

A variety of public databases and repositories grant researchers access to curated cancer datasets, including International Cancer Genome Consortium (ICGS), Gene Expression Omnibus (GEO), The cancer Genome Atlas (TCGA) [18]. These datasets facilitate comparative analyses, validation of research findings, and foster collaborative investigations [18].

Breast cancer prediction

The escalating annual mortality rates attributed to breast cancer, recognized as the most widespread form of cancer and a predominant contributor to female mortality on a global scale, underscore the pressing necessity for advancements in disease prognosis and detection methodologies to improve overall health outcomes. Achieving a high degree of precision in cancer forecasting is critically important for refining therapeutic approaches and augmenting patient survival statistics.

Machine Learning techniques are beneficial for detection of breast cancer and also enhance diagnostic efficiency and accuracy. Various studies have explored different ML techniques, demonstrating their potential to improve early detection and patient outcomes. machine learning algorithms such as SVM Classifier, LR, Random Forest, DT which examine the datasets to classify tumors as malignant or benign with high accuracy rates in detection of breast cancer [20]. In the present study, we use the Wisconsin Diagnostic Dataset for the purpose of breast cancer identification in this context, we perform experiment with various machine learning algorithms to find the accuracy of algorithm for breast cancer detection.

VI. RESULTS AND DISCUSSION

Data set description

The current study makes use of the Breast Cancer Wisconsin (Diagnostic) dataset, a well-known benchmark dataset to evaluate classification algorithms in the area of cancer diagnosis and healthcare data analysis. Dr. William H. Wolberg of the University of Wisconsin Hospitals in Madison, USA, was the original provider of the dataset. It consists of quantitative characteristics determined from digital pictures of fine needle aspirate (FNA) samples of breast masses. In order to identify whether a tumor is benign (non-cancerous) or malignant (cancerous), each example describes the properties of the cell nuclei visible in the picture.

Dataset Composition

The dataset comprises 569 samples and 31 attributes. Among these, 30 attributes are numeric features describing various properties of the cell nuclei, while the 31st column is the target variable indicating the diagnosis class. 30 attributes column describes numerical measurements of cell nuclei characteristics derived from fine needle aspirate (FNA) images of breast tissue. These include features such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. Each of these is computed as mean, standard error, and worst (maximum) values and target features describe Binary classification label representing the diagnosis of breast cancer: 0 denotes Malignant (cancerous), and 1 denotes Benign (non-cancerous). The features are grouped into three main categories:

1. Mean values – e.g., mean radius, mean texture, mean smoothness, mean symmetry, etc.
2. Error values – standard errors for each of the mean features (e.g., radius error, texture error, etc.).
3. Worst (maximum) values – maximum or "worst" values observed for each feature (e.g., worst radius, worst concavity, etc.).

Each sample therefore provides a comprehensive morphological description of a breast cell nucleus.

- The data set is split into 80:20 ratio of training and testing sets.
- The Standard Scaler is used to scale the features of the data set so that the mean is zero and the standard deviation is one. This improves the performance of machine learning algorithms that are sensitive to the size of the features.
- Various ML algorithm like SVM, KNN, Naïve Bayes, Logistic Regression, Random Forest, Decision Tree are trained using training data and their performance has been evaluated using testing data. The model's performance is measured using accuracy.
- Result shows most accurate model is Logistic Regression with accuracy 98.24 %.

Table II: Evaluation metrics for Machine Learning algorithm

Sr No	Machine Learning Algorithm	Accuracy	Precision	Recall	F1-Score
1	SVM Classifier	97.36 %	98.6 %	97.2%	97.9%
2	KNN Classifier	95.61 %	95.9%	97.2 %	96.5%
3	Naïve Bayes	92.98 %	94.4%	94.4%	94.4%
4.	Logistic Regression	98.24%	98.6%	98.6%	98.6%
5	Random Forest	95.61%	95.9%	97.2%	96.5%
6.	Decision Tree	91.22%	95.6 %	90.3%	92.9 %

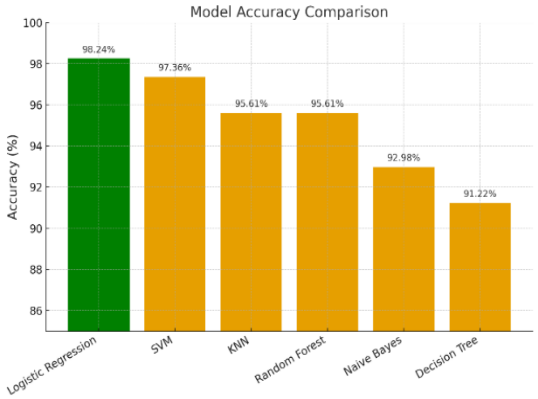

Fig 2: Comparison of Model accuracy for Brest cancer detection

VII CONCLUSION

Machine learning (ML) methodologies are indispensable across a multitude of business domains. The healthcare sector is encountering an increasing array of challenges, leading to a rise in associated costs. Numerous ML methodologies are employed to address these issues effectively. This manuscript elucidates a range of ML methodologies aimed at the prognostication of various medical conditions, including cardiac ailments, breast carcinoma, diabetes mellitus, and thyroid disorders. This review paper describes various techniques of machine learning for detection of chronical disease like heart disease, liver disease, breast cancer. Table II presents the performance evaluation of various machine learning approaches for breast cancer detection, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Logistic Regression, Random Forest, Decision Tree classifiers, using metrics such as accuracy, precision, recall, and F1-score. Fig 2 illustrates the comparative performance of different models, where logistic regression is highlighted in green color as it achieved highest accuracy of 98.24% compared to other model.

In the future, we aim to enhance the accuracy of breast cancer predictions by utilizing various deep learning algorithms and Explainable AI (XAI).

REFERENCES

[1] Sneha Grampurohit, Chetan Sagarnal, -Disease Prediction using Machine Learning Algorithms 2020, International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020

[2] Palle Pramod Reddy, Dirisinala Madhu Babu, Hardeep Kumar and Dr. Shivi Sharma, -Disease Prediction using Machine Learning" Sharma © 2021 IJCRT, Volume 9, Issue 5 May 2021, ISSN: 2320-2882

[3] Hafsa Habehh and Suril Gohel-Machine Learning in Healthcare, Department of Health Informatics, Rutgers University School of Health Professions, 65 Bergen Street, Newark, NJ 07107, USA. DOI: 10.2174/1389202922666210705124359

[4] Md Manjurul Ahsan, Shahana Akter Luna and Zahed Siddique, -Machine-Learning-Based Disease Diagnosis: A Comprehensive Review Healthcare 2022, 10, 541, DOI: 10.3390.

[5] K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, T. Suryawanshi -Human Disease Prediction using Machine Learning Techniques and Real-life Parameters, International Journal of Engineering. DOI: 10.5829/ije.2023.36.06c.07

[6] Taunk, K., De, S., Verma, S. and Sweta Padma, A., -A brief review of nearest neighbor algorithm for learning and classification, International Conference on Intelligent Computing and Control Systems (ICCS), IEEE. (2019), 1255-1260

[7] Dr. Nitish Das1, Sailee Gayke2, Nayan Patel3, Shrawani Shinde -Disease Prediction using Machine Learning, International Journal of Innovative Research in Science, Engineering and Technology, e-ISSN: 2319-8753, p-ISSN: 2347-6710 DOI:10.15680/IJIRSET.2024.1303276

[8] Ibrahim, I. and Abdulazeez, A., -The role of machine learning algorithms for diagnosing diseases, Journal of Applied Science and Technology Trends, Vol. 2, No. 01, (2021), 10-19. DOI: 10.38094/jastt2017.

[9] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain-A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms, International Journal of Scientific & Technology research, volume 8, Issue 11, November 2019 ISSN 2277-8616.

[10] Malti Bansal, Apoorva Goyal, Apoorva Choudhary, -A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short-Term Memory algorithms in machine learning, Analytics Journal 3 (2022) 100071, Published by Elsevier Inc.

[11] K. Shailaja B. Seetharamulu M. A. Jabbar, -Machine Learning in Healthcare: A Review, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

[12] Rakibul Islam1, Azrin Sultana1 and Mohammad Rashedul Islam, -A comprehensive review for chronic disease prediction using machine learning algorithms, Journal of Electrical Systems and Inf Technol (2024).

[13] Nur Ghaniaviyanto Ramadhan 1, Adiwijaya 2, (Member, IEEE), Warih Maharani 1, and Alfian Akbar Gozali, -Chronic Diseases Prediction Using Machine Learning with Data Preprocessing Handling: A Critical Review. IEEE Access 2024

[14] Yuan-Xing Liu a, Xi Liu a, Chao Cen a, Xin Li b, Ji-Min Liu c, Zhao-Yan Ming d, Song-Feng Yu a, Xiao-Feng Tang a, Lin Zhou a, Jun Yu a, Ke-Jie Huang b, Shu-Sen Zheng a ,-Comparison and development of advanced machine learning tools to predict non-alcoholic fatty liver disease: An extended study , Hepatobiliary Pancreat Dis Int 20(5):409–415

[15] Bo Zhang, Huiping Shi, Hongtao Wang, -Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach, Journal of Multidisciplinary Healthcare 2023:16 1779–1791

[16] Ahmad Ayid Ahmad, Huseyin Polat, -Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm, Diagnostics 2023,13,2392.

[17] Bo Zhang, Huiping Shi, Hongtao Wang, -Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach, Journal of Multidisciplinary Healthcare 2023:16 1779–1791

[18] Abrar Yaqoob, Rabia Musheer Aziz, Navneet Kumar verma, - Applications and Techniques of Machine Learning in Cancer Classification: A

Systematic Review, Human-Centric Intelligent Systems (2023) 3:588–615.

[19] Mahati Munikoti Srikantamurthy, V. P. Subramanyam Rallabandi, Dawood Babu Dudekula, Sathishkumar Natarajan, Junhyung Park, -Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning, BMC Medical Imaging, 2023 DOI: 10.1186/s12880-023-00964-0

[20] Vijeth G Poojari, -BREAST CANCER DETECTION USING MACHINE LEARNING International Research Journal of Modernization in Engineering Technology and Science, e-ISSN: 2582-5208, Volume:05/Issue:04/April-2023.

[21] David B. Olawade, Aanuoluwapo C. David-Olawade c, Ojima Z. Wada d, Akinsola J. Asaolu e, Temitope Adereni f, Jonathan Ling, -Artificial intelligence in healthcare delivery: Prospects and pitfalls, Journal of Medicine, Surgery, and Public Health 3 (2024) 100108.

[22] Xiaolu Tian and et al, -Using Machine Learning Algorithms to Predict Hepatitis B Surface Antigen Sero clearance, Hindawi Computational and Mathematical Methods in Medicine Volume, 2019

[23] A. Deva Kumar, Josephine Prem Kumar, V.S Prakash and Divya K.S, -Supervised Learning Algorithms: A Comparison, Kristu Jayanti Journal of Computational Sciences, Vol. 1, Issue 1, pp. 01-12, 2021

[24] Autsuo Higa, -Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms, International Journal of Computer Applications Technology and Research, Vol. 7, Issue 01, pp.23-27, 2018

[25] Kashvi Taunk, Sanjukta De, Srishti Verma Aleena, Swetapadma, -A Brief Review of Nearest Neighbor Algorithm for Learning and Classification, International Conference on Intelligent Computing and Control Systems (ICICCS 2019) IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8, DOI: 10.1109/ICCS45141.2019.9065747

[26] Shaik Anjum Niha1, Kolla Jhansi Lakshmi2, Pidathala Glory Bless3, Thamatam Sai Lakshmi4, Yarramasu Mounika Chowdary5, Mr. M. Purnachandra Rao6, -A Comparison of Machine Learning Algorithms for Predicting Liver Disease, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue III Mar 2023