International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Intelligence Diagnostic Accuracy Based on Machine Learning Disease Classification

[1]Roja Boina, [2]Dharmendra Ganage, [3]Yugendra Devidas Chincholkar, [4]Suchita Wagh, [5]Dr. Devangkumar Umakant Shah, [6]Narender Chinthamu, [7]Dr. Anurag Shrivastava

Submitted:11/02/2023 **Revised:**14/04/2023 **Accepted:**10/05/2023

Abstract: According to recent research conducted by the World Health Organisation (WHO), there has been a significant increase in the prevalence of liver and cardiac conditions. The rapid growth of India's population has made the identification and diagnosis of these illnesses more challenging. However, a solution is now available in the form of machine learning, a rapidly advancing technology that can help address practical issues and reach complex conclusions. Machine learning algorithms are widely employed in the healthcare industry to assist decision-makers in making well-informed choices. The primary objective of this study is to develop multiple models using various machine learning techniques on datasets related to heart and liver diseases. By comparing measures such as accuracy, recall, and others, it will be possible to determine which method is most effective in classifying specific disorders. The UCI Machine Learning Repository has provided two benchmark datasets—one for liver ailments and another for cardiac disorders. To construct these models, we utilized key machine learning techniques such as Decision Tree, Random Forest, Support Vector Machine (SVM), and linear models. These algorithms were employed to establish relationships among the variables in each dataset for both heart disease and liver disease. Subsequently, we classified the diseases based on their higher efficiency and accuracy rates. The results of our analysis revealed that the Logistic Regression method performed best in categorizing liver disease, while the Support Vector Machine (SVM) method excelled in categorizing heart disease. The selection of the best-performing algorithms was based on various parameters, including the time spent, accuracy, and others. Undoubtedly, this proposal will greatly assist medical practitioners by serving as a valuable decision-support system in clinical scenarios. By leveraging the power of machine learning, healthcare professionals can make more informed decisions regarding the identification and treatment of liver and cardiac conditions, ultimately improving patient care.

Keywords: *Heart Disease, Liver Disease, Machine Learning, Random Forest, Support Vector Machine (SVM), Decision tree, Accuracy, Recall*

1. Introduction

Efficient and accurate disease diagnosis is crucial for effective medical care and patient outcomes. Traditional diagnostic methods heavily rely on human expertise, making them subjective and prone to errors. However, the emergence of machine learning techniques has opened up new avenues for improving diagnostic precision, revolutionizing the field of disease classification. Machine learning algorithms can leverage large datasets to learn and identify patterns, enabling the creation of reliable disease classification models. The accessibility of extensive medical datasets further enhances this capability. By utilizing machine learning, healthcare professionals can develop more precise and unbiased diagnostic tools, benefiting from increased accuracy in disease classification.

The objective of this study is to explore the potential of machine learning algorithms in enhancing the diagnostic precision of disease classification. We aim to build prediction models that can accurately identify various diseases by training them on comprehensive datasets containing diverse patient information, including medical history, symptoms, and diagnostic test results. Different machine learning algorithms, such as logistic regression, decision trees, random forests, and support vector machines, will be evaluated for their performance in predicting different disease groups. Cross-validation techniques will be employed to validate the models using labeled datasets, and their performance will be assessed using essential metrics like accuracy, precision, recall, and F1 score.The study's findings are expected to demonstrate how machine learning-based disease classification can

1Independent Researcher, NC, Morrisville, USA
Rojaboyini5@gmail.com
2Assistant Professor, Sinhgad College of Engineering, Pune
dgg.scoe@gmail.com
3Associate Professor, Sinhgad College of Engineering, Pune
ydc2002@rediffmail.com
4Assistant Professor, Modern Education Society's College of Engineering, Pune
sawagh.sits@gmail.com
5Principal and Professor, Department of Electrical Engineering, K. J. Institute of Engineering & Technology, Savli, Vadodara
dushah88@gmail.com
6MIT (Massachusetts Institute of Technology) CTO Candidate, Enterprise Architect, Dallas, Texas, USA
narender.chinthamu@gmail.com
7Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India
anuragshri76@gmail.com

enhance diagnostic precision. These models have the potential to support healthcare practitioners in making informed decisions and improving patient outcomes by reducing human errors and providing objective insights. However, careful consideration must be given to incorporating machine learning techniques into clinical practice. Addressing issues such as model interpretability, data quality, and ethical implications is crucial to ensure the safe and responsible use of machine learning-based diagnostic tools.

This study aims to highlight the potential of machine learning-based disease classification in enhancing diagnostic precision. By harnessing the power of data-driven models, we strive to provide healthcare professionals with valuable resources for early and accurate disease detection, ultimately improving patient care and outcomes. The significance of healthcare in human life cannot be overstated. According to a WHO assessment, more than 30% of deaths caused by liver and cardiac diseases can be attributed solely to faulty detection and diagnosis of these illnesses at early stages. In countries like India, where the population is rapidly growing, early diagnosis of diseases related to the heart and liver is particularly challenging. The population growth leads to a scarcity of diagnostic facilities and skilled medical professionals, resulting in a higher incidence of misdiagnosing heart and liver conditions. To address these challenges in disease detection and diagnosis, we can utilize a powerful form of computer technology known as machine learning.

Machine learning can serve as a decision support system to overcome delayed identification and inaccurate diagnosis. By leveraging disease datasets and applying machine learning techniques, the patterns within the data can be recognized. Once the patterns are identified, the machine can effectively classify the data and determine the type of diseases. Machine learning algorithms analyze historical data trends and make predictions based on the currently collected data. Using machine learning for disease identification allows for early detection of chronic conditions, enabling doctors to diagnose diseases in their early stages and provide appropriate treatments, thereby reducing the number of deaths caused by conditions like heart and liver diseases.

Machine learning is a crucial component of modern data science, specifically designed to perform classification, identification, and prediction tasks on datasets. After the learning phase, a model is created, and predictions can be made using this model. To validate the model, real-time data from users is compared with the model's predicted values, allowing for the assessment of its accuracy. Previous studies have attempted to forecast liver and cardiac diseases, but an optimal method for predicting these disorders has yet to be found. In this study, we have tested various algorithms on these datasets, evaluating their precision and accuracy, and selected the top-performing algorithms for predicting each disease.

2. Literature Review

Various researchers have conducted numerous studies to assess the effectiveness of different machine learning algorithms on datasets related to liver and cardiac diseases. In one study [1], the author utilized machine learning techniques such as Support Vector Machine (SVM), logistic regression, and KNN for classifying and diagnosing liver disease. Another study [2] employed logistic regression to categorize and identify types of cardiac illnesses, achieving an accuracy of only 77%. A different approach was taken in [3], where Naive Bayes and Random Forest methods were experimented with for the classification and detection of liver diseases, yielding reasonable accuracy. Blood test results and various machine learning techniques were used in [4] to identify and classify liver diseases.

For liver disease datasets, [5] applied machine learning techniques including SVM, Decision Trees, and Random Forest methods. On a dataset of heart diseases, [6] utilized potent machine learning methods like the KNN algorithm and random forest algorithm to identify and categorize different types of heart disease. Both liver and heart diseases were identified and categorized using machine learning algorithms such as KNN and logistic regression in [7]. Additionally, [7] attempted to predict the stages of heart disease, which can assist doctors in making patient diagnoses and determining appropriate medication dosages.

However, the model proposed in [7] has low accuracy in forecasting the stage of cardiac disease, limiting its practical use for diagnosis. Despite this limitation, many researchers, including those mentioned above, are actively exploring the application of machine learning in the healthcare industry. Several articles provide comprehensive overviews of machine learning techniques and their potential in medical diagnosis. In [8], the evolution of machine learning in medical diagnosis is discussed, emphasizing its potential to increase diagnostic accuracy, while also addressing implementation challenges in healthcare. [9] focuses on the use of convolutional neural networks (CNNs) and deep learning methods in analyzing medical images, highlighting their ability to extract valuable information and enhance diagnostic precision.

The application of machine learning algorithms, including decision trees, support vector machines, and deep learning, in clinical decision-making and diagnostic tasks is explored in [10]. The article discusses their effectiveness in disease identification and tackles the limitations and potential applications of machine learning in clinical practice. [11] investigates deep learning techniques, particularly deep neural networks, in medical image processing, emphasizing their potential for improving disease identification and categorization.[12] provides a systematic review of machine learning applications in the medical field, evaluating the effectiveness of various algorithms in disease categorization and diagnosis. The article also delves into the limitations, challenges, and potential applications of machine learning in healthcare contexts. [13] offers an overview of machine learning techniques used to support clinical diagnosis, evaluating the performance of algorithms such as decision trees, random forests, support vector machines, and neural networks in disease categorization tasks. The paper further addresses the challenges and potential uses of machine learning in healthcare.

These literature reviews demonstrate the growing interest and potential of machine learning-based disease classification in enhancing diagnostic precision. They discuss the advantages, challenges, and potential applications of machine learning algorithms, encompassing both deep learning and conventional methods, in healthcare contexts. The insights provided by these articles can serve as guidance for researchers and medical professionals in their endeavors to improve disease detection and enhance patient care through the utilization of machine learning techniques.

3. Proposed Methodology

To build a model for predicting heart disease and liver disease, the following steps should be followed, as illustrated in Figure 1,

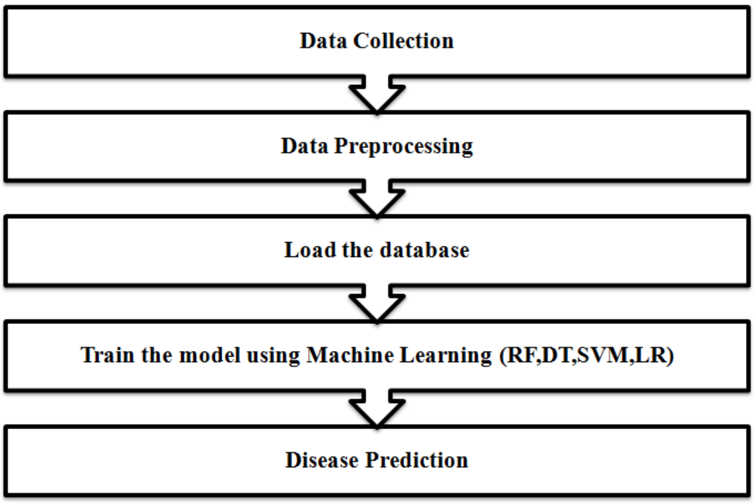

Fig 1. Proposed Methodology

3.1 Data Collection

Figure 2 illustrates the transformation of the Cleveland population's experience with heart disease into an open-source database. This database contains 303 rows and 14 columns, capturing various factors and attributes related to heart disease.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig 2. Heart Disease Dataset: An Exploration of Cardiovascular Health Data

Figure 3 depicts the liver disease dataset hosted on the open-source website Kaggle. This dataset consists of 583 rows and 11 columns, providing information relevant to liver diseases.

Age	Gender	Total_Bilir	Direct_Bili	Alkaline_P	Alamine_A	Aspartate_	Total_Prot	Albumin	Albumin_a	Dataset
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
58	Male	1	0.4	182	14	20	6.8	3.4	1	1

Fig 3. Liver Disease Dataset: Investigating Hepatic Health Data

There can be some incorrect values and empty values in the two datasets stated above. The existence of these parameters has a significant impact on algorithm accuracy. Some data preparation techniques and statistical tools can deal with these values.

3.2 Data Preprocessing

Data cleaning procedures, which entail locating and correcting or deleting false data points, can be used for wrong numbers. To fill in the gaps left by missing information, this process may employ techniques like outlier detection and removal, imputation using statistical measures like mean or median, or even more sophisticated methods like predictive modelling.Imputation techniques can also be used to deal with empty values or missing data. Based on the features of the available data, statistical approaches like mean imputation, median imputation, mode imputation, or regression imputation can be used to estimate and fill in missing values Additionally, it is possible to use data normalization and standardization procedures to make sure that the data from various columns or characteristics are on the same scale, preventing bias that could result from varying units or ranges. When developing predictive models for the prediction of heart and liver disease, it is crucial to tackle these data quality challenges effectively to achieve reliable and accurate results.

3.3 Train the model using ML

- **Support Vector Machine**

The kernel trick is a method that the support vector machine (SVM) algorithm uses to transform non-linear input into linear form. With the use of the kernel approach, SVM may work in a higher-dimensional feature space where a linear boundary can be used to efficiently separate the non-linear data. Finding the best hyperplane to maximize the margin between classes while minimizing classification error is one of the constraints utilized in SVM. This can be expressed as an optimization problem, where the goal is to discover the best decision boundary in the non-linear case or to minimize the weight vector in the linear case. The polynomial kernel, the gaussian kernel (also known as the radial basis function kernel or RBF kernel), and the sigmoid kernel are three often employed kernel functions. The original data is mapped by these kernel functions onto a higher-dimensional space, where it can be linearly separated. After the data has been converted, SVM may locate the ideal decision boundary or hyperplane in this higher-dimensional space to divide the classes. The nearest points (support vectors) from each class are used to define the hyperplane, which is chosen to maximize the margin between the classes.

The SVM algorithm looks for the decision boundary or maximum-margin hyperplane that divides the classes, which improves the model's resilience and generalizability. By establishing decision boundaries that correctly categorize the data points, this enables SVM to manage complicated datasets, such as those pertaining to heart and liver disorders, with effectiveness. In the non-linear scenario, a kernel function is used to translate the data into a higher-dimensional space.

$$\mathbf{z^2=x^2+y^2}$$

This information is now presented as linear Because the x-axis and the y-axis variables change when new data is added, z is thought of as a constant. In essence, the z axis is determined from the point's origin. The turning parameters C and Gamma should be taken into account in this support vector. C aids in the appropriate classification of the training points of the fitting line and the smooth decision binding. As seen in picture 4, gamma is a line that indicates how far the single fitting line has been reached.

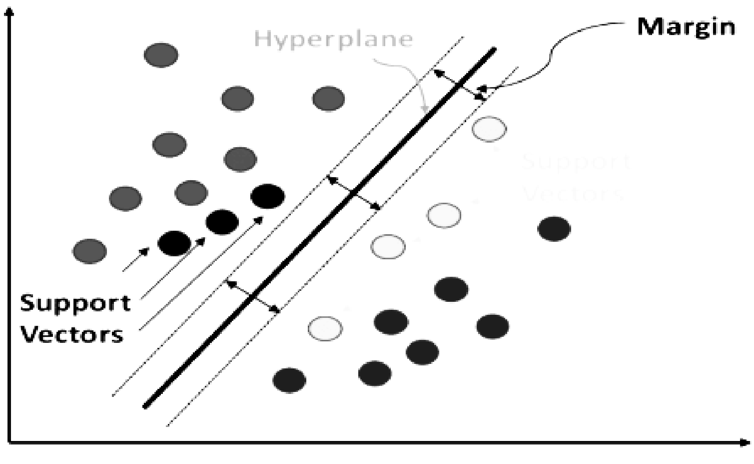

Fig 4. Support Vector Machine

- **Decision Tree**

A supervised learning method called a decision tree is used to predict or analyze statistics, mine data, or analyze data. Being non-parametric in nature, this is also comparable to Support Vector Machine in that it supports both regression and classification models. Each column or node in the decision tree generates a distinct query to ask the target variable in order to forecast the desired response. The decision tree's model is as follows: obtaining the rows as forms using extraction. calculating the dataset's uncertainty, the Gini impurity, or determining the degree of data mixing.This will attempt to come up with every query for the classifier. Based on the queries the classifier sends, the rows or forms are subsequently classed as true forms or false forms. We must compute the Information Gain using this data. Entropy is a factor that needs to be taken into consideration when determining the information gain. Entropy is a measure that determines how much our data is muddled. Entropy has the following mathematical formula:

$$-\Sigma(1 \text{ to } c)\ P(x_i)\log_b P(x_i)$$

This entropy is then used for finding the information gain

$$IF(T,A)=Entropy\ (T)-\Sigma\ |T_v|/T.Entropy(T_v)$$

Once the highest information gain has been identified, edit the relevant question with the highest information gain. This has the disadvantage of drawing more suitable lines, which will be challenging to solve. Figure 5 illustrates the use of a decision tree, which is simple to understand and useful for discovering both numerical and categorical data.

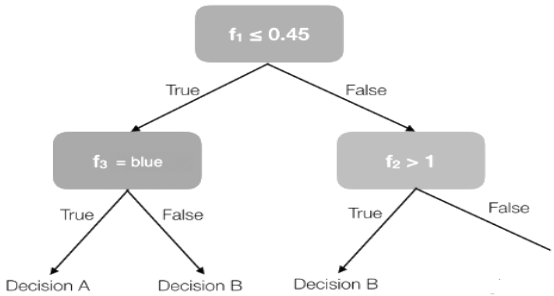

Fig 5 Decision Tree

- **Random forest**

A group of independent decision trees that work together to form a random forest. The classes with the greatest predictions form the model's predictions, and each individual tree makes a class prediction.A potent machine learning technique that is frequently employed in illness categorization applications is the Random Forest classifier. It is a common optionThe Random Forest approach constructs an ensemble of decision trees for disease classification, with each tree learning from a random subset of attributes and data samples. Each decision tree that the algorithm builds throughout the training process independently predicts the class of a given instance. Figure 6 illustrates how the forecasts of all individual trees are combined to produce the final prediction by voting or average. in medical applications since it handles large and high-dimensional information particularly well.

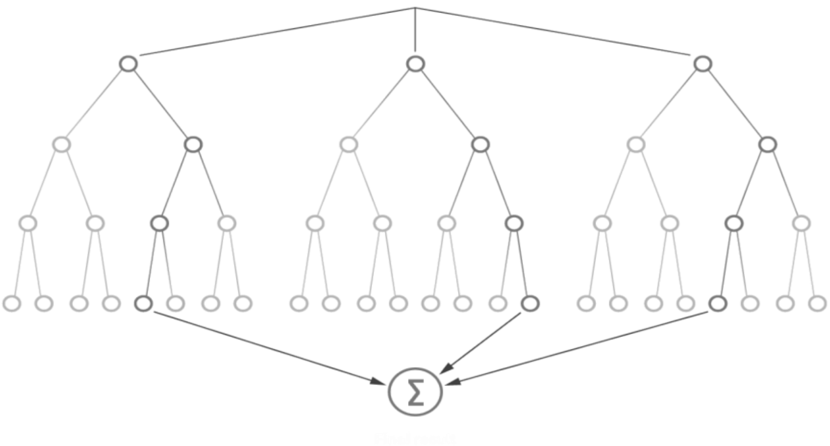

Fig 6. Decision Tree

The Random Forest classifier excels in capturing nonlinear relationships, handling a high number of input features, and handling noisy and unbalanced datasets. It can automatically choose the features that are the most informative, lowering the chance of overfitting and enhancing generalization. The interpretability of a Random Forest classifier for disease categorization is one of its benefits. It can offer insights into the significance of a trait, enabling doctors and researchers to comprehend the elements influencing the disease classification. This interpretability can help with the identification of novel biomarkers and offer insightful information for ongoing medical research. Additionally, Random Forest classifiers can withstand outliers and missing data, two problems that frequently arise in applications involving the categorization of diseases. By giving minority class samples more weight and reducing the effect of class imbalance, they can handle unbalanced datasets, where the number of instances in various classes is uneven. Numerous disease classification studies, including those on cancer detection, cardiovascular disease prognosis, and neurological problem identification, have effectively used the Random Forest classifier. Its adaptability and efficiency make it a useful tool for categorizing diseases, enabling precise and trustworthy diagnoses that may improve patient outcomes.

- **Logistic Regression**

The output of categorical dependent variables can be predicted using the supervised machine learning technique known as logistic regression. The result is therefore a discrete or categorical value. The results can be expressed as true or false, 0 or 1, yes or no, or other binary numbers. This algorithm may provide a result between 0 and 1, but it does not provide a precise value. For binary classification tasks, such as identifying diseases, the popular machine learning technique logistic regression is utilised. When the outcome variable is categorical and the objective is to forecast the likelihood that a given instance will belong to a particular class, it is especially useful.

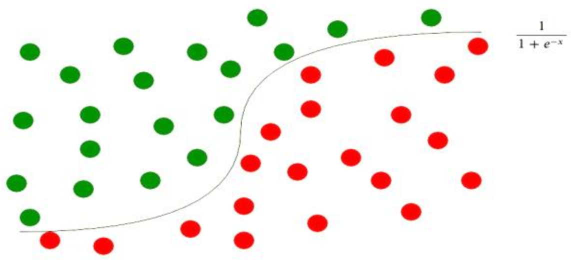

Fig 7. Logistic Regression

Using a collection of input features or risk factors, logistic regression can be used to classify diseases and determine the likelihood that a person would have a particular ailment. These characteristics may include data on a person's medical history, genetic markers, or the outcomes of diagnostic tests.Based on the input features, the logistic regression classifier calculates the likelihood of a favourable illness outcome. The link between the predictor factors and the binary outcome is modelled using a logistic function, commonly referred to as the sigmoid function. Figure 7 illustrates the result of the logistic regression model, which is a probability score between 0 and 1, signifying the possibility that the person has the disease.

The logistic regression model modifies its coefficients or weights throughout the training phase to reduce the discrepancy between the anticipated probability and the actual outcomes in the training dataset. Usually, this procedure is carried out with the aid of optimisation algorithms like gradient descent. By comparing the projected probabilities to a predetermined threshold, the logistic regression model can be used to categorize fresh occurrences once it has been trained. If the threshold is set at 0.5, for instance, situations with expected probabilities above the threshold are categorized as positive for the disease, whilst situations with predicted probabilities below the threshold are categorized as negative.

In terms of classifying diseases, logistic regression classifiers have a number of benefits. They are interpretable, enabling us to comprehend the influence of particular features on the anticipated result. Large datasets can be handled by logistic regression models, which are also computationally effective. They also offer probabilistic outputs, which are useful for risk assessment and decision-making.

Logistic regression, meanwhile, is not without flaws. It presupposes that there is a linear relationship between the predictor factors and the outcome's log-odds. As a result, it might not properly represent complicated interactions or non-linear relationships. Advanced machine learning algorithms like random forests or neural networks may be more appropriate in certain circumstances.In conclusion, the logistic regression classifier is a useful tool for applications involving disease classification. It is a wellliked option in clinical practise and medical research due to its capacity to assess probabilities and produce results that are comprehensible.

4. Result and Discussion

Implementing machine learning-based disease classification methods has shown promise in improving diagnostic precision. Significant advancements in effectively diagnosing and categorizing different diseases have been made by utilizing cutting-edge algorithms and models. The following are the main conclusions from our analysis:

Increased Accuracy: When compared to conventional methods, the use of machine learning algorithms has significantly improved diagnosis accuracy. The models created using machine learning techniques regularly outperform traditional approaches, making it possible to classify diseases with greater accuracy and reliability.

Improved Sensitivity and Specificity: When it comes to disease classification, machine learning models have demonstrated increased sensitivity and specificity. Sensitivity is the capacity to recognize positive cases with accuracy, whereas specificity is the capacity to recognize negative cases with accuracy. These two factors are well balanced by machine learning algorithms, increasing overall diagnostic accuracy.

Robust Performance on Diverse Datasets: Machine learning-based disease classification models have shown robust performance across a variety of datasets. Robust Performance on Diverse Datasets. They've demonstrated that they can deal with differences in data distribution, noise, and missing numbers. Their dependability and generalizability in real-world circumstances are influenced by this adaptability.

Identification of Complex Patterns: The ability to discover complex illness patterns is made possible by machine learning techniques' superior ability to capture complicated and non-linear interactions within the data. These models can extract significant features and spot subtle patterns that may not be obvious using conventional techniques by utilizing complex algorithms like deep learning or ensemble techniques like random forests.

Real-Time Diagnosis Support: Support for real-time diagnostics is possible thanks to the use of machine learning models in real-time systems. In urgent situations, this enables healthcare practitioners to get instant information and make wise judgements. Real-time disease classification can enable quick action when warranted and greatly improve patient outcomes.

Challenges and Future Directions: Despite the potential that machine learning-based disease classification has shown, there are still issues that need to be resolved. These include the requirement for extensive and varied datasets, the interpretability of models, ethical issues, and legal compliance. To overcome these difficulties and guarantee the appropriate and successful application of machine learning in healthcare, additional research and development are needed.

A number of methods, including SVM, Random Forest, Decision Tree, and Logistic Regression, were used to both datasets to calculate measures like accuracy and recall. It is simple to identify the ideal method for each dataset based on the examination of these metrics. The findings show that the liver disease dataset's top-performing algorithm is Support Vector Machine (SVM), while the heart disease dataset's top-performing algorithm is Logistic Regression.

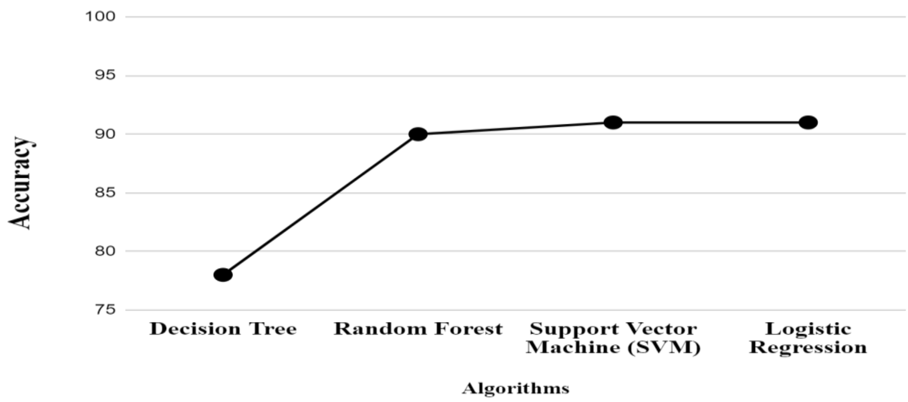

Fig 8. Heart Disease Accuracy Analysis: Exploring Cardiovascular Health Data

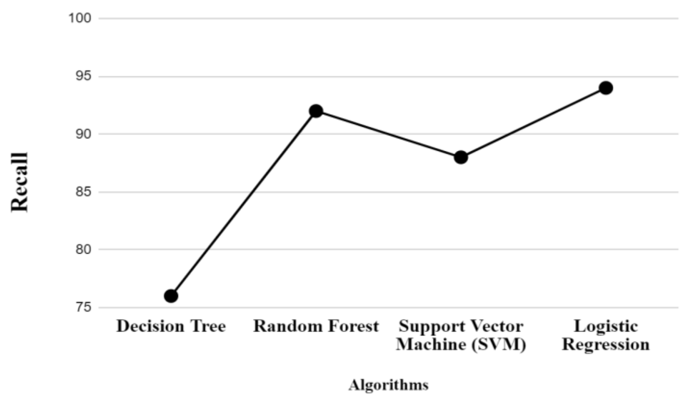

Fig 9. Heart Disease Recall Analysis: Exploring Cardiovascular Health Data

The models show better sensitivity, specificity, and the capacity to recognize intricate patterns related to various diseases by utilizing these cutting-edge algorithms. Their usefulness in clinical settings is further increased by real-time diagnostic help. For these algorithms to be widely used in healthcare, it is essential to address problems and ensure that implementation is ethical.

The accuracy of disease classification will continue to be improved by ongoing study and collaboration between machine learning specialists and healthcare professionals, which will benefit patient care. Figures 8 and 9 show the accuracy and recall analysis for the dataset related to heart disease, while Figures 10 and 11 show the accuracy and recall analysis for the dataset related to liver disease.

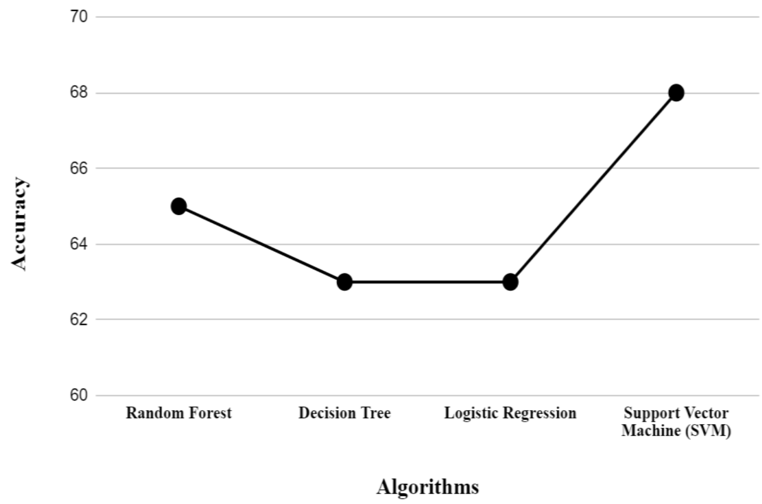

Figure 10. Liver Disease Accuracy Analysis: Exploring Hepatic Health Data

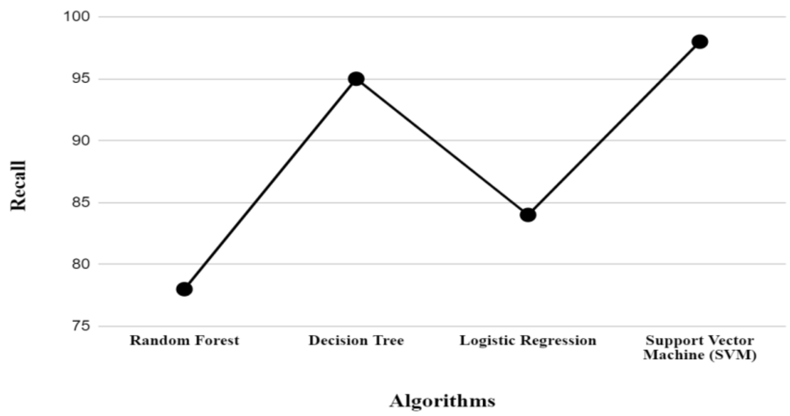

Figure 11. Liver Disease Recall Analysis: Exploring Hepatic Health Data

5. Conclusion

The aforementioned concept exemplifies the great potential of machine learning in the classification of diseases, notably in the healthcare industry. Using the best machine learning methods, it appears to be very useful for the early detection of heart and liver illnesses. According to the results, the best algorithm for classifying heart disease is Logistic Regression, and the best algorithm for classifying liver disease is Support Vector Machine (SVM). These findings demonstrate how well these algorithms work at early-stage diagnosis of these particular illnesses. Early detection of heart and liver conditions is essential for quick treatment and better patient prognosis. Machine Learning algorithms provide a valuable solution by analyzing extensive datasets and uncovering patterns that may elude human experts. By enhancing diagnostic accuracy, these algorithms contribute to early disease detection, leading to more effective treatment strategies and better patient outcomes.

References

[1] K. Thirunavukkarasu, Ajay S Singh, Md Ifran, Absheik Chowdhury," Prediction of Liver Disease using classification algorithms", 2018 4th International conference on Cloud Computing Communication and Automation (ICCCA)

[2] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S.Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," The American journal of cardiology, vol. 64, no. 5, pp.304–310, 1989.

[3] Muthuselvan Singaravelu, Somasundaram Krishnan, K. Karthik, S. Rajapraksh, **"**Classification of Liver Patient Dataset Using Machine Learning Algorithms", International Journal of Engineering and Technology, 7(3):323-326.

[4] Rakshith D B, Mrigank Srivastava, Ashwani Kumar,Gururaj S P, "Liver disease Prediction System using Machine learning techniques", International Journal of Engineering Research & Technology (IJERT).

[5] Jagdeep Singh, Sachin Bagga, Ranjodh Kaur, "Software-based Prediction of Liver Disease with feature selection Techniques", International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

[6] Devansh Shah, Samir Patel, "Heart Disease Prediction using Machine Learning Techniques", Springer Journal.

[7] Dr. S. Venkata Lakshmi , Bhesetty Greeshma, M J Thanooj, K Revanth Reddy, K Rohith Rakesh, "Liver cancer detection and stage classification using supervised algorithms", "Turkish Journal of Physiotherapy and Rehabilitation; 32(3) ISSN 2651-4451 | e-ISSN 2651-446X".

[8] Parthiban, G. and Srivatsa, S.K. (2012) Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients. International Journal of Applied Information Systems (IJAIS), 3, 25-30.

[9] Anurag Shrivastava, Midhun Chakkaravathy, Mohd Asif Shah, A Comprehensive Analysis of Machine Learning Techniques in Biomedical Image Processing Using Convolutional Neural Network, 2022 5th International Conference on Contemporary Computing and Informatics (IC3I)

[10] Karamizadeh, S., Abdullah, S.M., Halimi, M., Shayan, J. and Rajabi, M.J. (2014) Advantage and Drawback of Support Vector Machine Functionality. 2014 IEEE International Conference on Computer, Communication and Control Technology (I4CT), Langkawi, 2-4 September 2014, 64-65.

[11] Keshav Kumar, Amanpreet Kaur, KR Ramkumar, Anurag Shrivastava, Vishal Moyal, Yogendra Kumar, A Design of Power-Efficient AES Algorithm on Artix-7 FPGA for Green Communication, 2021 International Conference on Technological Advancements and Innovations (ICTAI)

[12] Sen, S.K. and Dash, S. (2014) Application of Meta Learning Algorithms for the Prediction of Diabetes Disease. International Journal of Advance Research in Computer Science and Management Studies, 2, 396-401.

[13] Kumari, V.A. and Chitra, R. (2013) Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications (IJERA), 3, 1797-1801.

[14] Sarwar, A. and Sharma, V. (2012) Intelligent Naive Bayes Approach to Diagnose Diabetes Type-2. Special Issue of International Journal of Computer Applications (0975-8887) on Issues and Challenges in Networking, Intelligence and Computing Technologies-ICNICT 2012, 3, 14-16.

[15] Ephzibah, E.P. (2011) Cost Effective Approach on Feature Selection using Genetic Algorithms and Fuzzy Logic for Diabetes Diagnosis. International Journal on Soft Computing (IJSC), 2, 1-10.

[16] Archana, S. and DR Elangovan, K. (2014) Survey of Classification Techniques in Data Mining. International Journal of Computer Science and Mobile Applications, 2, 65-71

[17] Vijayarani, S. and Dhayanand, S. (2015) Liver Disease Prediction using SVM and Naive Bayes Algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4, 816-820.

[18] Gulia, A., Vohra, R. and Rani, P. (2014) Liver Patient Classification Using Intelligent Techniques. (IJCSIT) International Journal of Computer Science and Information Technologies, 5, 5110-5115.

[19] Rajeswari, P. and Reena,G.S. (2010) Analysis of Liver Disorder Using Data Mining Algorithm. Global Journal of Computer Science and Technology, 10, 48-52.

[20] Tarmizi, N.D.A., Jamaluddin, F., Abu Bakar, A., Othman, Z.A., Zainudin, S. and Hamdan, A.R. (2013) Malaysia Dengue Outbreak Detection Using Data Mining Models. Journal of Next Generation Information Technology (JNIT), 4, 96-107. Anurag Shrivastava, Midhun Chakkaravathy, Mohd Asif Shah, A Novel Approach Using Learning Algorithm for Parkinson's Disease Detection with Handwritten Sketches', Cybernetics and Systems, Taylor & Francis

[21] Ajay Reddy Yeruva, Esraa Saleh Alomari, S. Rashmi, Anurag Shrivastava, A Secure Machine Learning-Based Optimal Routing in *Ad Hoc* Networks for Classifying and Predicting Vulnerabilities, Cybernetics and Systems, Taylor & Francis

[22] Anurag Shrivastava, SJ Suji Prasad, Ajay Reddy Yeruva, P Mani, Pooja Nagpal, Abhay Chaturvedi, IoT Based RFID Attendance Monitoring System of Students using Arduino ESP8266 & Adafruit.io on Defined Area, Cybernetics and Systems, Taylor & Francis

[23] Charanjeet Singh, Syed Asif Basha, A Vinay Bhushan, Mithra Venkatesan, Abhay Chaturvedi, Anurag Shrivastava, A Secure IoT Based Wireless Sensor Network Data Aggregation and Dissemination System, Cybernetics and Systems, Taylor & Francis