International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

Volume 10, Issue 7, May-June-2024 | Published Online : 20th June 2024

Machine Learning-Based Disorder Prediction System

Keerthana R, Pratibha Prakash Machakannur

UG student, Department of Computer Science and Engineering New Horizon College of Engineering, Bangalore, India

***Abstract*— This study addresses the limitation of existing healthcare machine learning models, which primarily focus on detecting single diseases, by developing a system capable of predicting multiple diseases through a unified interface. The proposed model aims to forecast diseases such as diabetes, chronic kidney disease, and cancer such as Melanoma, all of which carry significant risks if left untreated. Early detection and diagnosis facilitated by this system could potentially save lives. The research employs various classification algorithms including K-Nearest Neighbour, Support Vector Machine, Decision Tree, Random Forest, Logistic Regression, and Gaussian Naive Bayes, comparing their accuracies to identify the most effective predictor. Multiple datasets specific to each disease are utilized to ensure the highest level of accuracy. The primary goal is to build a web application capable of predicting multiple diseases, including diabetes, chronic kidney disease, and cancer such as Melanoma using machine learning.**

Keywords: Unified interface, Diabetes, Chronic kidney disease, Cancer Melanoma diagnosis, K Nearest Neighbor, Support Vector Machine, Convolutional neural networks (CNN), Decision Tree, Random Forest, Logistic Regression, Medical imaging, Early detection, Feature extraction, Retinal blood vessel

I. INTRODUCTION

The objective of this research is to forecast several ailments, encompassing diabetes, heart disease, chronic kidney disease, and cancer. A range of classification methodologies, including KNN, SVM, Random Forest, Logistic Regression, and CNN, are deployed for disease prognosis. The accuracy of each technique is verified and juxtaposed to identify the optimal predictor. Multiple datasets tailored to each disease are utilized to augment prediction precision. The top-performing algorithm for each ailment is chosen and incorporated into a web-based platform. This platform allows users to input disease-specific parameters for convenient prediction of the desired ailmentThis study utilizes various datasets, each specific to a separate disease.

For cardiac disease, datasets from the Cleveland, Hungary, Switzerland, and Long Beach V databases are employed. These datasets contain 76 attributes, with 14 selected for analysis. The "target" field indicates the presence of heart disease, with 0 denoting no disease and 1 denoting the presence of disease.

The dataset for chronic kidney disease consists of 25 features and was collected over 2 months in India. Attributes such as red blood cell count and white blood cell count are included. The classification is binary, with "ckd" indicating chronic kidney disease and "notched" indicating its absence. This dataset comprises 400 records.

The dataset for diabetes includes records from the Pima Indians Diabetes Database and a Kaggle dataset. It consists of one target variable, "Outcome," and multiple medical predictor factors such as BMI, insulin level, and age. This dataset comprises 769 records and 9 columns.

After collecting datasets from various sources, data pre-processing techniques such as label encoding are applied. Models are then created using various machine learning algorithms, including K-NN, Gaussian NB, Decision Trees, Support Vector Machine, Logistic Regression, CNN and Random Forest.

Each disease dataset is split into training and testing sets, and each model is trained on the training dataset. The accuracy of each model is evaluated against the testing dataset, and the best-performing model is selected.

The web application's user interface features a sidebar for navigation and forms for entering input attribute values for a particular disease. The sidebar is created using the option_menu method of the streamlit_option_menu module, while the input value fields in the forms are generated using the text_input method of the streamlet.

The trained models (classifiers) are easily loaded into the Stream lit editor for prediction using the pickle module.

II. LITERATURE SURVEY

D. Roja Ramani [1] The paper presents a novel method to enhance melanoma diagnosis by combining segmentation and 3D feature extraction. It introduces techniques such as multi-atlas segmentation and patch-based label fusion to improve accuracy. By reconstructing 3D skin lesions from depth maps and analyzing streaks' characteristics, the system achieves better classification. Additionally, utilizing Deep Convolutional Neural Networks enhances segmentation and classification accuracy. The paper's contributions include 3D image reconstruction, streak line detection, and the application of advanced models for segmentation. Overall, it offers a comprehensive approach to melanoma diagnosis, addressing key challenges and improving efficiency in skin cancer detection.

T. Jemima Jebaseeli, [2] The paper proposes a framework to enhance retinal blood vessel segmentation in diabetic retinopathy patients using adaptive histogram equalization

and a pulse-coupled neural network (PCNN) model. Unlike previous methods, this approach dynamically selects the PCNN's threshold limit based on seed points, enabling one-time segmentation. It also employs a bidirectional searching procedure and standardized neuron weights for improved efficiency. Evaluation on fundus image datasets shows enhanced segmentation accuracy in terms of sensitivity, specificity, and accuracy.

Himanshu Kriplani [3] The paper introduces a DNN technique for CKD prediction, attaining 97% accuracy on 224 records. It outperforms random forest and support vector machine algorithms. Preprocessing and cross-validation are employed to prevent overfitting. Early CKD detection is vital for mitigating health and financial burdens. The study underscores the efficacy of advanced ML methods like DNNs in this domain.

Arumugam, K [4] The Cleveland data set is used. After cleaning the data using by preprocessing, Machine learning algorithms like SVM, Nave Bayes, and Decision Tree C4.5 are now fed this data. These classifiers are used to predict heart disease in diabetes individuals. Mohit.

Inductura [5] In this endeavor, a web application has been created to identify diseases such as breast cancer, diabetes, and heart disease by employing machine learning models like logistic regression, SVM, and K-Nearest Neighbors.

KM Jyothi Rani [6] Diabetes dataset used in this work contains 2000 cases. Predicting whether the person is diabetic or not is the objective. Different classification algorithms used are KNN, Logistic Regression, Decision Tree, Random Forest, and SVM.

Dr. Sunita Varma [7] Goal of this work is to predict diabetes with better accuracy. Various classification and ensemble algorithms like Random Forest, SVM, KNN, Decision Tree, Logistic Regression, and Gradient Boosting classifiers are used to predict diabetes. The Pima Indian Diabetes Dataset repository at UCI is where the information was found.

Quan Zou [8] Data from hospital physical examinations in Luzhou, China was used to create the dataset. Diabetes is predicted using machine learning classification techniques including decision trees, random forests (RF), and neural networks. The Luzhou dataset results demonstrate the shortcomings of the blood glucose-free approaches.

Nazid Ahmed and Gazi Mohammed Ifraz [9] they tells the Preprocessing is performed on datasets, and various machine learning algorithms are used for prediction, followed by deployment of the best model in a web application using Flask. The focus is on predicting diabetes, with considerations for model performance metrics and deployment.

[10] In this work, three different models were trained for accurate prediction using a variety of physiological variables as well as ML methods as logistic regression (LR), decision tree (DT) classification, and Knearest neighbor (KNN). Only kidney illness is predicted by the system's design. Accuracy can be raised by using a larger, deeper dataset with more attributes.

S.Revathy [11] Machine learning prediction algorithms can be used intelligently to anticipate the onset of CKD and provide a means of early treatment. This study suggests the best prediction model when attempting to predict CKD utilizing classifiers like Decision Trees, Random Forests, and Support Vector Machines. Only kidney illness is predicted by the system's design. Accuracy can be raised by using a larger, deeper dataset with more attributes.

Zixian Wang [12] Based on a dataset for chronic kidney disease (CKD) from the UCI machine learning data warehouse, this study analyses chronic kidney disease using machine learning techniques. For 400 individuals with chronic kidney disease, the A priori association approach is used to identify CKD. Only kidney illness is predicted by the system's design. Accuracy can be raised by using a larger, deeper dataset with more attributes.

F.J Shaikh [13] The ML & DL techniques utilized in cancer progression modeling are proposed to be reviewed in this research. Many of the predictions discussed are associated with certain ML, input, and data sample supervision. can employ additional cutting-edge machine learning algorithms and extraction techniques to provide a more thorough comparison analysis.

Baban Uttamrao [14] The suggested approach compares the outcomes of HRFLM application to other classification methods, such as decision trees and support vector machines. One dataset is used by the developed system, which only predicts heart disease. In some instances, it causes inaccuracies.

A. Requirement analysis

1) FUNCTIONAL REQUIREMENTS

Disorder Prediction Functionality:
Implement machine learning algorithms (Naïve Bayes, K-NN, Random Forest, Logistic Regression, SVM) to predict multiple diseases based on input attributes.

Model Training and Validation:
Train and validate each algorithm's accuracy against the datasets to find the best predictor for each disease.

Integration of Multiple Datasets:
Integrate datasets specific to heart disease, cancer, diabetes, and chronic kidney disease for accurate prediction.

Web Application Development: Develop a user-friendly web application using Streamlit to allow users to input disease-specific attributes and obtain predictions.
Testing and Evaluation:
Test the web application's functionality, including the prediction process, and evaluate its performance against predefined metrics.

2) NON-FUNCTIONAL REQUIREMENTS

Performance:

Ensure that the system can handle a large amount of data efficiently and provide timely predictions.

Usability:
Design the web application interface to be intuitive and easy to use for users with varying levels of technical expertise.

Reliability:
Ensure that the prediction models are reliable and accurate, providing consistent results across different datasets.

III. METHODOLOGY

Fig.1 The process of predicting multiple diseases using a range of machine learning algorithms, such as Naïve Bayes, K-NN, Random Forest, Logistic Regression, and SVM, is outlined. The aim is to facilitate effective communication between patients and doctors, enabling both parties to achieve their respective goals efficiently. Each algorithm's accuracy is verified and compared to determine the most reliable predictor. Incorporating multiple datasets enhances the precision of predictions. Furthermore, a user-friendly web application has been created to streamline the disease prediction process. Users can input specific attribute values for the disease they wish to predict, simplifying the entire procedure.

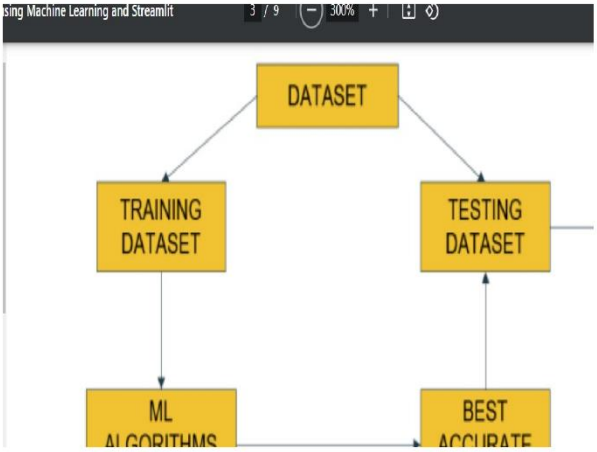


Fig. 1. System Architecture of Proposed Work

The advantages of the proposed system include:

Prediction of multiple diseases utilizing various machine learning techniques.

Efficient analysis of large datasets within a minimal timeframe.

Validation and comparison of algorithmic accuracies to identify the optimal prediction method.

A. melanoma

1) methodology for melanoma detection using machine learning along with equations where applicable:

1. Data Collection: Gather a dataset of images containing both melanoma and non-melanoma (benign) skin lesions. This dataset should be labeled accordingly.

2. Data Preprocessing: Preprocess the images to standardize them for the model. This may include resizing, normalization, and augmentation techniques to increase the diversity of the dataset and improve model generalization.

3. Feature Extraction: Use techniques like Convolutional Neural Networks (CNNs) to automatically extract relevant features from the images. CNNs are particularly well-suited for image classification tasks due to their ability to capture spatial hierarchies of features.

4. Model Training: Train a machine learning model using the pre-processed images and their corresponding labels. One common approach is to use transfer learning, where a pre-trained CNN model (e.g., VGG, Reset, Inception) is fine-tuned on the melanoma detection dataset. This process involves freezing certain layers of the pre-trained model and retraining only the final layers with the new dataset.

5. Model Evaluation: Evaluate the trained model on a separate test dataset to assess its performance. Common evaluation metrics for binary classification tasks like melanoma detection include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC).

6. Deployment: Once the model achieves satisfactory performance, deploy it in a production environment. This could involve creating a web application using a framework like Streamlit, Flask, or Django, where users can upload images of skin lesions for melanoma detection.

2) Equations:

1. Convolution Operation:

$$C[i,j] = \sum_m \sum_n I[i+m, j+n] \cdot K[m,n]$$

Where:

$$C[i,j] = \sum_m \sum_n I[i+m, j+n] \cdot I$$

Where:

- $C[i,j]$ is the output pixel value at loc

2. Activation Function (e.g., RELU):

$f(x) = \max(0,x)$

This function introduces non-linearity into the network and helps capture complex patterns in the data.

3. Loss Function (e.g., Binary Cross-Entropy):

This function introduces non-linearity into

Where:

data.

3. Loss Function (e.g., Binary Cros

B. Diabetic detection

Detecting diabetes using machine learning involves training a model on a dataset containing features relevant to diabetes diagnosis, such as glucose levels, blood pressure, BMI, etc. Streamlit is a useful tool for building interactive web applications, including those for machine learning models. Below, I'll provide a simplified methodology, code snippets, and equations to give you a starting point. For the sake of brevity, I'll focus on a basic logistic regression model and a simple Streamlit web app

1) Methodology:

Data Collection: Gather a dataset containing features related to diabetes diagnosis and their corresponding labels (diabetic or not).
Data Preprocessing: Clean the data, handle missing values, normalize/standardize features, etc.
Model Selection: Choose a suitable machine learning algorithm (e.g., logistic regression, decision trees, etc.).
Model Evaluation: Evaluate the model's performance using metrics like accuracy, precision, recall, etc.
Building Streamlit Web App: Develop an interactive web application using Streamlit to deploy the trained model.

Sure, let's delve deeper into the algorithm and equation for logistic regression, which is commonly used in binary classification tasks like diabetic detection.

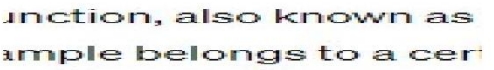
2) Algorithm: Logistic Regression

Logistic regression is a statistical method utilized to forecast the result of a categorical dependent variable by considering one or more predictor variables. It finds extensive application in tasks involving binary classification.

Here's the algorithm for logistic regression:
1. Initialize Parameters: Initialize the weights (coefficients) and the bias term (intercept) randomly or with some pre-defined values.
2. Compute Linear Combination: For each data point, compute the linear combination of the input features and their corresponding weights, along with the bias term.
3. Apply Sigmoid Function: Pass the linear combination through the sigmoid (logistic) function to squash the output between 0 and 1, interpreting it as the probability of the positive class.
4. Compute Loss: Compute the loss between the predicted probabilities and the actual labels using a suitable loss function, such as binary cross-entropy loss.
5. Update Parameters: Update the weights and bias using gradient descent or another optimization algorithm to minimize the loss.
6. Repeat: Repeat steps 2-5 until convergence or for a fixed number of iterations.

Equation: Logistic Function (Sigmoid Function)

The logistic function, also known as the sigmoid function, is used to give input samples belonging to a certain class.
It's defined as:

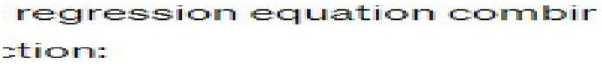


Where:
(z) The linear combination of input features, their corresponding weights, and the bias term comprise the logistic regression model.
(e) is the base of the natural logarithm.

The logistic function maps the output of the linear combination to the range [0, 1], representing the probability of the positive class. If the probability is greater than or equal to 0.5, the model predicts the positive class; otherwise, it predicts the negative class.

Logistic Regression Equation

The logistic regression equation integrates the linear combination of input features with the logistic function.



Where:
(P(y=1|x)) is the probability that the target variable (y) (e.g., diabetic or not) is 1 (positive class) given the input features (x).
(w) is the vector of weights (coefficients) for each feature.
(x) represents the vector of input features.
(b) dotes the bias term, also known as the intercept.

The model predicts the positive class (1) if (P(y=1|x)) is greater than or equal to 0.5, and the negative class (0) otherwise.

This equation encapsulates the essence of logistic regression, where the model learns the relationship between the input features and the probability of belonging to a certain class. Training involves finding the optimal values for the weights (w) and the bias term (b) to minimize the loss function.

C. Chronic Kidney Diseases

The study utilized the chronic kidney disease dataset from the UCI Machine Learning Repository. This dataset was uploaded in 2015 and contains data collected from the Apollo Hospital in Tamil Nadu over nearly two months. The dataset comprises 25 attributes, including 11 numeric and 14 nominal features.
In the study, a total of 400 instances from the dataset were used. Among these instances, 224 were utilized for training prediction algorithms. Within this training set, there were 105 instances labeled as chronic kidney disease and 119 instances labeled as non-chronic kidney disease.

1) Optimization

Gradient descent is a technique used for minimizing the cost function in machine learning training epochs. When the cost function is convex, gradient descent follows the slope to reach the minimum. However, for non-convex functions, stochastic gradient descent is employed, which stochastically finds local minima. To address this, adaptive techniques like Adam adjust learning rates independently for each parameter. Finally, K-fold cross-validation is

utilized to partition the dataset into K subsets for training and testing the model.

2) Algorithm:

Data Preprocessing:

- Clean the dataset (handling missing values, outliers).
- Encode categorical variables.
- Scale numerical features.

Feature Selection:

- Identify relevant features using correlation analysis or domain knowledge.

Split Data:

- Split the dataset into training and testing sets.

Deployment:

- Deploy the model in a Streamlit application for CKD detection.

Logistic Regression Equation:

Logistic regression predicts the probability of a binary outcome, such as a patient having CKD, based on input features. The logistic regression equation represents this relationship succinctly.

outcome. In our case, it predicts the probability of a patient ha

The logistic regression equation is as follows:

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n)}}$$

Where:

This equation calculates the probability of a patient having CKD (Y=1) based on the input features X and model coefficients β. The logistic function (sigmoid function) ensures that the output stays between 0 and 1, representing probabilities.

In the Streamlit application, the equation is used to calculate the probability of CKD for each patient, and a threshold is applied to classify patients into CKD-positive or CKD-negative categories based on their probabilities.

3) Result and Discussion

The model, incorporating 18 parameters and various optimization techniques, achieves an accuracy of 97.76% and an F1 measure of 97.6%, as indicated in Table 1. Other algorithms applied to the dataset were also evaluated based on True Positive Rate (TPR), False Positive Rate (FPR), Precision, Recall, and F-measure.

Precision signifies the proportion of correctly classified positive instances among all instances labeled as positive, whereas Recall indicates the percentage of positive instances correctly classified by the model..

Our artificial neural network exhibits superior Precision, Recall, and True Positive Rate compared to other algorithms, with Naïve Bayes showing similar Precision. Specifically, the neural network identifies chronic kidney disease instances with a True Positive Rate of 95.2%. Further training with more data may enhance accuracy and speed, without requiring human intervention. Refer to Table 1 for a detailed comparison of accuracy and performance.

Table 1 Accuracy and performance by class

		True positive	False positive	Precisic
Chronic kidney disease	Naïve Bayes	0.952	0	1
	Deep neural network	0.952	0	1
	Logistic	0.943	0	1
	Random forest	0.952	0	1
	Adaboost	0.962	0	1
	SVM	0.962	0	1
Nonchronic kidney disease	Naïve Bayes	1	0.048	0.96
	Deep neural network	1	0.048	0.96

Prediction of Chronic Kidney Diseases Using Deep Artificial

Table 2 Kappa statistics, roc and rmse

	Root mean squared error	Accuracy (%)	Ka
Naïve Bayes	0.1407	97.7679	0.9
Deep neural	0.1214	97.7679	0.(

Table 2 presents the accuracy of all algorithms along with kappa statistics, ROC (Receiver Operating Characteristic) curve, and root mean square error. The graph in Figure 4 illustrates the decrease in model loss with an increase in epochs. The loss function tends towards zero for the training set and slightly above zero for the testing set.

Root means square error provides a magnitude of error regardless of its direction, while kappa statistics indicate the percentage of agreement, with a score of 0.95 signifying near-perfect agreement in predicting chronic kidney disease. The ROC curve represents the area under the curve, with a value of 1 indicating almost full coverage, as observed in our neural network model.

186

Table 3 Confusion matrix for the comparable method

	NOT chronic kidney diseas
Naïve Bayes	119(TP)
	5(FP)
Deep neural network	119(TP)
	5(FP)
Logistic	119(TP)
	6(FP)
Random forest	119(TP)
	2(FP)

Table 3 displays the confusion matrices of all implemented algorithms. In the case of the deep neural network, there were 119 true positive predictions and 100 true negative predictions. Notably, the test set used for predictions was distinct from the training dataset, with correct classifications into chronic and non-chronic kidney disease outlined in the table.

In our neural network model, among 224 records, 219 were accurately classified (true negatives and true positives). Naïve Bayes presents a similar confusion matrix but exhibits a higher root mean squared error compared to our model, potentially due to zero probability estimates for certain attribute values. Logistic regression demonstrates a higher false positive rate and doesn't encompass 100% of the area under the ROC curve, necessitating a sizable sample size for stability. Random forest performs admirably with a reduced false negative rate, whereas AdaBoost achieves slightly lower coverage under the ROC curve but is sensitive to noise and inclined towards overfitting. Support vector machine showcases good performance but demands meticulous parameter tuning and kernel selection, thereby transferring the challenge of overfitting to model selection.

D. Machine Learning Models

1) Gaussian Naïve Bayes

Gaussian Naive Bayes is a simple method for creating classifiers, which uses models to assign issue instance class labels that are given as vectors of factor values and are selected from a finite set. There isn't a single technique for training these classifiers, but rather a tribe of algorithms founded on a common tenet: given the class variable, all naïve Bayes classifiers infer that a certain feature's value is independent of any other feature's value.

2) K Nearest Neighbor

The K-nearest Neighbors algorithm is a versatile supervised learning technique utilized for both classification and regression purposes. It operates by categorizing new data based on its similarity to existing data points, employing the concept of grouping. By measuring the resemblance between the new and existing data, it assigns the new data to a category that closely matches its characteristics. This algorithm is popularly used for classification tasks, relying on the assumption that neighboring points with similar features tend to belong to the same category.

3) Support Vector Machine

SV-Machine is a procedural methodology that is being used for classification & regression challenges. On the other hand, this algorithm is specifically utilized for classification issues in real-world problems and scenarios. Each data point is represented using this method as a point in n-D space, where n is the total number of features, each of which has a unique feature. To identify the hyper-plane that separates the classes in the supplied data set, the classification technique is then called.

4) Decision Tree Both

classification and regression issues can be solved using decision trees. It is a greedy algorithm or ID3. Only a question is posed in a decision tree, and subtrees are created based on the answer (yes/no). Recursively, input data is divided up based on chosen properties. The attribute selection measures are used to determine the best attribute for the root node and sub-node. Since it uses the same reasoning process that people use when making decisions, it is easy to understand.

5) Random Forest

(RF) works by creating various decision trees on different subsets of data and the outputs of those individual decision trees are used together to form a final output. Random forest works on the bagging principle, first, various subsets are generated from the dataset this process is called bootstrap. Now each model works on these subsets and gives outputs, next random forest combines the outputs of various models and produces output based on voting this step is known as aggregation.

E. Streamlit

Elegant machine-learning web apps may be swiftly produced and distributed using an open-source framework called Streamlit. It is a Python library designed primarily with machine learning users in mind. Since the majority of data scientists are familiar with this framework, their work has been made easier because they aren't interested in devoting weeks to learning how to utilize it to build web applications. The good thing about Streamlit is that it allows you to create your own online application right away, even if you don't have any previous experience with web development. Therefore, Streamlit is a great choice if you're interested in data science and want to deploy your models quickly, simply, and with the least amount of code possible.

Features of Streamit:

- No requirement for HTML, CSS, or JavaScript.
- Rapid development of impressive machine learning or data science software within hours or even minutes, instead of days or months.
- Compatibility with a wide range of Python libraries, including Pandas, Matplotlib, Seaborn, Plotly, Keras, PyTorch, and SymPy (latex).
- Creation of amazing online applications with minimal code.
- Simplified and accelerated computation pipelines through data caching.

CONCLUSION

In, this research endeavors to forecast and facilitate the prediction of ailments including diabetes, chronic kidney

disease, and cancer as melanoma. Various classification methodologies such as KNN, SVM, Random Forest, Logistic Regression, and CNN were explored and juxtaposed to identify the most effective predictor for each ailment. Utilizing disease-specific datasets, encompassing attributes relevant to each condition, the study deployed rigorous machine-learning techniques for model training and evaluation. The chosen top-performing algorithms were then integrated into a user-friendly web-based platform, allowing convenient input of disease-specific parameters for prediction. Notably, for cardiac disease, datasets from multiple sources were amalgamated, while chronic kidney disease data was collected over two months in India, and diabetes records were sourced from established databases. Through meticulous data preprocessing and model training, the platform offers accurate and accessible predictions tailored to individual ailments. This interdisciplinary approach serves as a valuable tool in healthcare decision-making and management, potentially aiding early detection and intervention in critical medical conditions Further enhancements and validations can be explored to continually refine the predictive capabilities of the platform and extend its utility in real-world healthcare.

REFERENCES

[1] D. Roja Ramani , & S. Siva Ranjani.” An Efficient Melanoma Diagnosis Approach Using Integrated HMF Multi-Atlas Map Based Segmentation” Journal of Medical Systems (2019) 43:225.

[2] Himanshu Kriplani, Bhumi Pate Research in Computer and Communication Engineering 8.12 (2019): 50-52.

[3] Alanazi, Rayan. "Identification and prediction of chronic diseases using machine learning approach." Journal of Healthcare Engineering 2022 (2022).

[4] Arumugam, K., et al. "Multiple disease prediction using Machine learning algorithms." Materials Today: Proceedings (2021). [5] Mohit, Indukuri, et al. "An Approach to detect multiple diseases using machine learning algorithm." Journal of Physics: Conference Series. Vol. 2089. No. 1. IOP Publishing, 2021.

[6] Mujumdar, Aishwarya, and V. Vaidehi. "Diabetes prediction using machine learning algorithms." Procedia Computer Science 165 (2019): 292-299.

[7] Joshi, Tejas N., and P. P. M. Chawan. "Diabetes prediction using machine learning techniques." Ijera 8.1 (2018): 9-13.

[8] Zou, Quan, et al. "Predicting diabetes mellitus with machine learning techniques." Frontiers in genetics 9 (2018): 515.

[9] Ahmed, Nazin, et al. "Machine learning based diabetes prediction and development of smart web application." International Journal of Cognitive Computing in Engineering 2 (2021): 229- 241.

[10] Ifraz, Gazi Mohammed, et al. "Comparative analysis for prediction of kidney disease using intelligent machine learning methods." Computational and Mathematical Methods in Medicine 2021 (2021).

[11] Revathi, S., et al. "Chronic kidney disease prediction using machine learning models." International Journal of Engineering and Advanced Technology 9.1 (2019): 6364-6367.

[12] Wang, Zixian, et al. "Machine learning-based prediction system for chronic kidney disease using associative classification technique." International Journal of Engineering & Technology 7.4.36 (2018): 1161.

[13] Shaikh, F. J., and D. S. Rao. "Prediction of cancer disease using machine learning approach." Materials Today: Proceedings 50 (2022): 40-47.

[14] Rindhe, Baban U., et al. "Heart Disease Prediction Using Machine Learning." Heart Disease 5.1 (2021).