Journal of Computer Science

Review Article

IntelliHealth: A Machine Learning Driven Disease Detection and Diet recommendation System

[1]Ankita Wadhwan, [2]Priyanka Chawla, [1]Sandeep Kaur and [1]Usha Mittal

[1]*School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*
[2]*Department of Computer Science and Engineering, National Institute of Technology, Warangal, India*

Article history
Received: 12-07-2024
Revised: 16-10-2024
Accepted: 29-10-2024

Corresponding Author:
Priyanka Chawla
Department of Computer Science and Engineering, National Institute of Technology, Warangal, India
Email:
priyankachawla.cse@gmail.com

Abstract: People all around the world are afflicted with various ailments. An accurate diagnosis can lower the risk of significant health problems developing, but an inaccurate diagnosis could have adverse implications. In this study, an ensemble-based strategy "IntelliHealth" has been presented to identify disorders of the thyroid, liver, and breast cancer using three machine learning (ML) approaches consisting of Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). The datasets for this research work are acquired from Kaggle. The experimental results show that the ML based ensemble model provides the highest level of disease prediction accuracy. This model is 93% accurate for liver, 99% accurate for breast cancer, and 100% accurate for diabetes and thyroid. Also, in this study, a web-based application is developed that uses proposed ensemble for quickly predicting diseases based on the patient's profile and recommends a diet plan.

Keywords: Healthcare, Real Time Disease Prediction, Diet Recommendation, Ensemble, Classification

Introduction

Due to the widespread usage of technology nowadays, academics are focusing more on creating web-based systems that employ a Machine Learning (ML) method to give those systems additional intelligence. The use of websites benefits healthcare workers in many ways; it enhances the whole healthcare system and promotes efficient decision-making. Cutting-edge healthcare applications may significantly decrease the amount of time patients spend waiting for appointments, consultations, and diagnostics. In this study, ML models are applied to predict liver disease, diabetes, and thyroid and breast cancer. The body's largest organ is the liver. Numerous conditions, including viral infections, heavy drug usage, poisoning, alcoholism, obesity, and many more, can result in various liver problems. Since the liver can still operate normally even while partially infected, early liver disease is sometimes difficult to diagnose. This makes it challenging for doctors to provide accurate early projections. Early diagnosis and prognosis can enhance liver healing and stop significant issues from emerging.

Diabetes is one of the chronic illnesses characterized by high sugar (glucose) levels in the blood. In this study, the Pima Indian Diabetes (PIDD) Dataset has been utilized to train and validate the ML models. Thyroid is also a very common disease, which is typically caused by iodine deficiency. An endocrine gland that delivers hormones into the bloodstream is the thyroid gland. The thyroid gland produces hormones that promote hydration, balanced body functions, and proper digestion. Cancer is the second most common cause of mortality globally. Women are most frequently killed by breast cancer. If certain symptoms develop, breast cancer is often detectable early. However, a large number of breast cancer patients show no symptoms. For an early diagnosis, routine breast cancer screening is crucial. Large-scale medical datasets for disease prediction are a problem that is addressed using ML approaches.

Users of the proposed web application merely need to provide a few details, like gender, Body Mass Index (BMI) and age and the outcome is shown right away. The patient's health status and symptoms must be assessed in order to provide an accurate diagnosis of the ailment. The proposed method can more precisely identify any condition and is not type-dependent. With a precise and speedy diagnosis, the fatality rate can be lowered. Additionally, this online healthcare platform offers diet programs depending on the patient's profile.

The working of a health recommender system is shown in Figure (1). The various datasets from which data can be obtained are described by data management. ML techniques are used to predict the diseases by training models on collected datasets. It also suggests a diet plan based on the patient's profile.

SCIENCE Publications

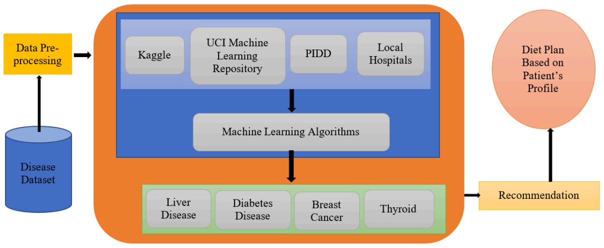


Fig. 1: Recommendation system

In this study, IntelliHealth, an ensemble-based machine learning system has been designed to detect thyroid, liver and breast cancer diseases and provide personalized diet recommendations. The proposed system leverages Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF) algorithms to achieve high accuracy in disease prediction. The integration of diet recommendations is based on a comprehensive analysis of patient profiles, including age, gender, BMI and specific symptoms.

The major work done in the proposed work are as follows:

- Data prepossessing approaches are applied on the datasets of Diabetes, Liver, breast cancer and thyroid disease to clean and standardize them
- After data prepossessing, different models are trained with the machine learning algorithms such as SVM, RF and DT
- Hyper parameter tuning is performed to select best model
- An ensemble model is designed using the models selected after hyper parameter tuning
- A Web application is designed to help laymen users to access the proposed model

Literature Review

Various disease prediction systems have been developed by using ML and DL algorithms. It includes predictions for diseases like liver disease, thyroid, breast cancer and diabetes. In this section, related works for diagnosis of various diseases have been summarized.

Diabetes

Krishnamoorthi *et al.* developed ML based system based on PIDD dataset. The data was analyzed using bar graphs. The outliers were removed and standardization was performed on the data. The implementation method's results demonstrate how the Linear Regression (LR) outperformed other ML algorithms. The ROC value of LR has been discovered to be 86% (Krishnamoorthi *et al.*, 2022). Sultana *et al.* applied multiple linear regression technique. Using this model, 83% accuracy has been attained. The data was gathered from various websites where more than 1,000 users reported their real-time health indices. Diet recommendation was provided using reinforcement-based algorithm (Sultana *et al.*). Ramesh *et al.* presented that the accuracy, sensitivity and specificity performance measures for this study were able to get scores of 83.20, 87.20 and 79%, respectively, by the ten-fold stratified cross-validation approach. Imputation was performed to account for missing data, followed by feature scaling to standardize the range of the dataset values. Feature selection methods were applied to remove redundant features that did not contribute significantly to the prediction outcome, enhancing overall model fidelity. To rectify class imbalances, an oversampling method was used to synthesize similar samples of the minority class during the data augmentation step (Ramesh *et al.*, 2021). Alaa Khaleel and Al-Bakry (2023) implemented three ML classifiers that is K-Nearest Neighbor (KNN), LR and Naïve Bayes (NB) and achieved the accuracy of 69, 94 and 79% and respectively. According to the findings, LR is superior to other algorithms at predicting diabetes Alaa Khaleel and Al-Bakry (2023). Rajput *et al.* (2022) determined the best ML model, the prediction were performed using LR, SVM, Random Forest (RF), Decision Tree (DT), NB, KNN, Kernel Approximation algorithm (KA) and Stochastic Gradient Descent algorithm (SGD). SVM is the algorithm of choice with the highest accuracy of 96.0%. The effectiveness of the ML model is assessed using a variety of test methods, including 10-fold cross-validation and achieved the accuracy of 66% (Bhat and Ansari (2021). Nadeem *et al.* predicted the disease using SVM, Artificial Neural Network (ANN) Fusion and SVM-ANN, with SVM having an accuracy of 88.30%, ANN Fusion of 93.63 and SVM-ANN having an accuracy of 94.67%. Two different datasets were collected and fused to train the models. Datasets were preprocessed by replacing null values with mean and feature scaling was done using standardization. Model selection was performed using K fold cross validation where K was taken as Choudhury and Gupta (2019); (Nadeem *et al.*, 2021). Choudhury and Gupta, used DT, LR, KNN, NB and SVM to forecast the illness. It has been discovered that LR provides the most accurate results for classifying samples with and without diabetes. These algorithms were put to use on the PIMA Indian Diabetic dataset, which has a total of 768 samples and 9 characteristics (Choudhury and Gupta (2019). Sarwar *et al.* applied SVM, KNN, LR, DT, RF and NB to predict diabetes using 8 features and 768 samples in the PIMA Indian dataset. 77% accuracy is the maximum provided by SVM and KNN (Sarwar *et al.*, 2018). Alehegn *et al.* employed 768 recorded diabetes datasets from UCI and the SVM, Bayes Net, decision stump and Proposed Ensemble Method (PEM) algorithms to predict the disease. The decision stump provided less accuracy than other methods, with an accuracy value of 83.72%, while the proposed technique had a high accuracy value of 90.36% (Alehegn *et al.*, 2018). Dey *et al.* represented various ML techniques on the Indian Pima dataset, ANN offers us the maximum accuracy (82.35%) using the min-max scaling method (Dey *et al.*, 2018). Table (1) shows the summary of the diabetes disease prediction using various algorithms.

Table 1: Summarized review of diabetes disease recommendation systems

Author	Dataset	Technique/Algorithm	Accuracy %
Krishnamoorthi *et al.* (2022)	Pima Indian Diabetes Database	LR	20
		KNN	64
		SVM	75
		RF	84
		Proposed LR	91
Sultana *et al.* (2021)	Different web sources	MLR	83
Ramesh *et al.* (2021)	University of California, Irvine ML repository	KNN	79.80
		LR	73.30
		GNB	73.10
		SVM-RBF	83.20
Alaa Khaleel and Al-Bakry (2023)	PIDD	KNN	69
		NB	79
		LR	94
Rajput *et al.* (2022)	PIDD	LR	95
		SVM	96
		RF	79
		DT	89
		NB	85
		KNN	83
Bhat and Ansari (2021)	Local hospital	DT	87
		NB	90
		RF	93
Nadeem *et al.* (2021)	PIDD	SVM	88.30
		ANN Fusion	93.63
Choudhury and Gupta (2019)	PIDD	DT	67
		SVM	75
		KNN	75
		NB	76
		LR	77
Sarwar *et al.* (2018)	(UCI) MLR held by the National Institutes of Diabetes	LR	74
		SVM	77
		NB	74
		RF	71
		DT	71
		KNN	77
Alehegn *et al.* (2018)	PIDD	SVM	88.8
		DS	83.72
		PEM	90.36
Dey *et al.* (2018)	PIDD	SVM	78.05
		KNN	75.5
		GNB	79.3
		ANN	82.35

Figure (2) shows the percentage of work on diabetes disease using various methodologies. It has been observed that a maximum amount of work has been done by SVM and ensemble models.

Thyroid

Rehman *et al.* analyzed of various classifiers, including FS techniques implemented in KNN, NB, SVM, DT and LR. The outcome showed that classifiers using L1-based FS outperformed classifiers using the L2-based FS techniques in terms of overall accuracy NB100, LR100 and KNN 97.84% (Abbad Ur Rehman *et al.*, 2021a).

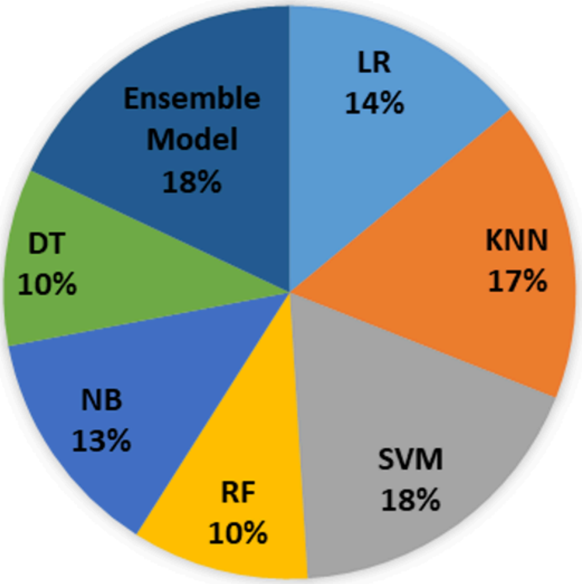


Fig. 2: Percentage of work done using different techniques on diabetes

Shivastuti *et al.* implemented SVM and RF techniques to predict thyroid illness. The data preprocessing has been performed by converting categorical values into numerical values and for normalization min-max normalization has been used. SVM performed better than RF with accuracy, precision and recall of 93, 89 and 93%, respectively. SVM was discovered to be a more successful method for predicting thyroid disorders (Shivastuti *et al.*, 2021). Vasile *et al.* collected photos of 230 thyroid disease patients, with a higher female gender predominance. Dataset consists of 30 children between the ages of 17 days-17 years and 200 adults between the ages of 18-75. 78. 75 individuals had hypothyroidism confirmed, 15 had subclinical hypothyroidism, 2 had congenital hypothyroidism, 38 had hyperthyroidism confirmed, 10 had preclinical hyperthyroidism and 90 had euthyroidism. The collected dataset was augmented using vertical and horizontal translations, random shearing and random zooming transformations. Further, data preprocessing has been performed using image cropping and image resizing. It has been observed from the results that the ensembled method outperformed with the accuracy of 97.35% (Vasile *et al.*, 2021). Chaubey *et al.* represented three commonly used ML algorithms, LR, DT and KNN. They compared and contrasted these algorithms in order to predict and assess their performance in terms of accuracy. The thyroid data set from the ML repository in the UCI knowledge discovery database archive has been used for this. KNN 96.87% achieved a higher accuracy as a result (Chaubey *et al.*, 2021). Rehman *et al.*

presented the application of KNN with its many distance functions, such as Euclidean, Manhattan, Canberra, Cosine and Minkowski is shown to identify thyroid illnesses. With the use of the chi-square-based FS technique for the newly proposed dataset, the performance of the three different distance functions-KNN, Euclidean and Cosine-achieved 100% accuracy at the k value of 1 (Abbad Ur Rehman *et al.*, 2021b). Raghuraman *et al.* combined SVM, MLR and DT ML algorithms. Data preprocessing has been implemented using mean imputation technique. The SVM classification model had the lowest accuracy, while the DT model had the highest accuracy at 97.35% (Raghuraman *et al.*, 2019). Raisinghani *et al.* applied many techniques such as DT, RF, SVM, ANN and LR, for the construction of thyroid illness prediction models. The accuracy of the DT algorithm is found to be 99.46%, giving it a higher score (Raisinghani *et al.*, 2019). Tyagi *et al.* gave examples of various ML methods and thyroid preventive diagnoses. With an accuracy of SVM 99.63, KNN 98.62, DT 75.76 and ANN 97.50%, ML algorithms, SVM, KNN and DT, were used to predict the estimated risk of a patient's likelihood of developing thyroid illness (Tyagi *et al.*, 2018). Shankar *et al.* compared the suggested thyroid classification to the current model, the findings are 97.49, 99.05 and 94.5% accuracy, sensitivity and specificity, respectively (Shankar *et al.*, 2020).

The summary of the different techniques of Thyroid disease prediction is given in Table (2).

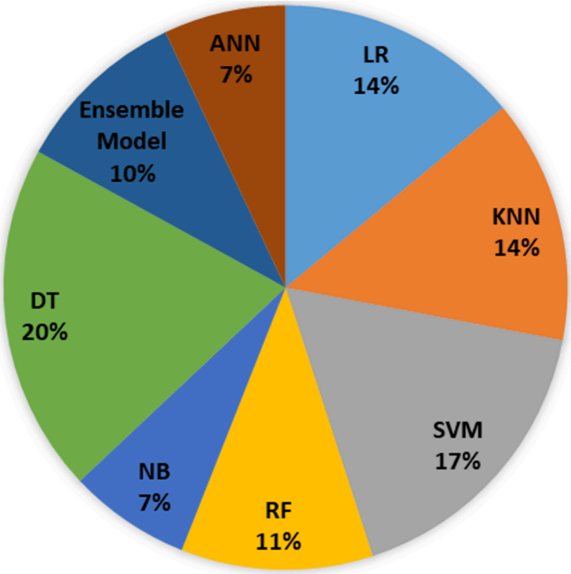


Fig. 3: Percentage of work done using different techniques on Thyroid Disease

Figure (3) gives the percentage of work on thyroid disease detection using various methodologies. It has been observed that Decision Tree is maximum utilized for thyroid disease detection.

Table 2: Summary of thyroid disease recommendation systems

Author	Dataset	Technique/Algorithm	Accuracy %
Abbad Ur Rehman *et al.* (2021a)	DHQ teaching hospital	KNN	97
		DT	75.34
		NB	100
		SVM	86.02
		LR	100
Shivastuti *et al.* (2021)	Thyroid Dataset, from the UCI	SVM	93
		RF	92
Vasile *et al.* (2021)	Institutions of Craiova	CNN-VGG	97.35
Chaubey *et al.* (2021)	Thyroid data set from UCI	DT	87.5
		KNN	96.87
		LR	81.25
Abbad Ur Rehman *et al.* (2021b)	District Headquarters teaching hospital, Dera Ghazi Khan, Pakistan	Euclidean	100
		Cosine Minkowski	99.3
		Manhattan Canberra	98
		Shebyshev correlation	77.3
Raghuraman *et al.* (2019)	thyroid dataset from the (UCI) MLR	DT	97.35
Raisinghani *et al.* (2019)	thyroid disease dataset from (UCI) MLR	DT	99.46
		RF	99.30
		SVM	96.25
		LR	97.50
		ANN	95.17
Tyagi *et al.* (2018)	thyroid disease dataset from (UCI) MLR	SVM	99.63
		KNN	98.62
		DT	75.76
		ANN	97.50
Shankar *et al.* (2020)	Thyroid dataset from UCI	MK-SVM	97.49

Breast Cancer

Allugunti (2022) employed Convolutional Neural Networks (CNN), SVM and RF algorithms to forecast breast cancer. A total of about 1000 images, representing roughly 150 patients with or without breast cancer, were included in the dataset. Data augmentation was applied to increase the size of the dataset. The images were resized and object-oriented picture segmentation was applied to remove salt and pepper noise. Features such as border, thickness, colour etc. were extracted from the images. Different models were trained out of which CNN provided the highest accuracy of 99.67 percent whereas SVM's accuracy was 89.84 percent and RF's accuracy of

90.55 percent. Mridha (2021) implemented KNN, SVM, RF, NB, LR, GB, ANN, etc., to the dataset used. The accuracy of the RF solution is 98.83%, whereas the KNN algorithm accuracy is 91.22% (Mridha, 2021). Islam *et al.* compared SVM, KNN, RF, ANN and LR five supervised ML methods on the dataset was taken from the UCI repository. The maximum accuracy was achieved by ANNs 98.57% (Islam *et al.*, 2020). Ak (2020) implemented ML methods like LR, KNN, SVM, NB, DT, RF and rotation forest. The LR model's results revealed a maximum classification accuracy of 98.1% with all characteristics included (Ak, 2020). Sharma *et al.* utilized RF, KNN and NB for predicting breast cancer (Sharma *et al.*, 2018). Yarabarla *et al.* employed methodologies including KNN, SVM, RF and GB to forecast the disease. GB achieved the highest accuracy of 73% (Yarabarla *et al.*, 2019). Omondiagbe *et al.* developed a hybrid method for diagnosing breast cancer by first utilizing SVM on the new reduced feature dataset after using LDA to lower the high dimensionality of the features. CFS, RFE, PCA and LDA are the approaches taken into account. The datasets were trained using ANN, NBC and SVM (Omondiagbe *et al.*, 2019). Tapak *et al.* presented a dataset containing the records of 550 breast cancer patients, the following models were used to predict breast cancer survival and metastasis: NB, RF, AdaBoost, SVM, Least-Square SVM (LSSVM), Adabag, LR and LDA. 83.4 percent of the 550 patients were still alive, and 85% did not develop metastases (Tapak *et al.*, 2019). Tahmooresi *et al.* used numerous ML techniques, including SVM, ANN, KNN and DT for efficient breast cancer detection. According to these researchers, SVM is the most frequently applied technique for cancer detection applications. The highest accuracy possible with SVM (single or hybrid) was 99.8% and it can be increased to 100% (Tahmooresi *et al.*, 2018). Table (3) provides the summary of the Breast cancer prediction using various algorithms.

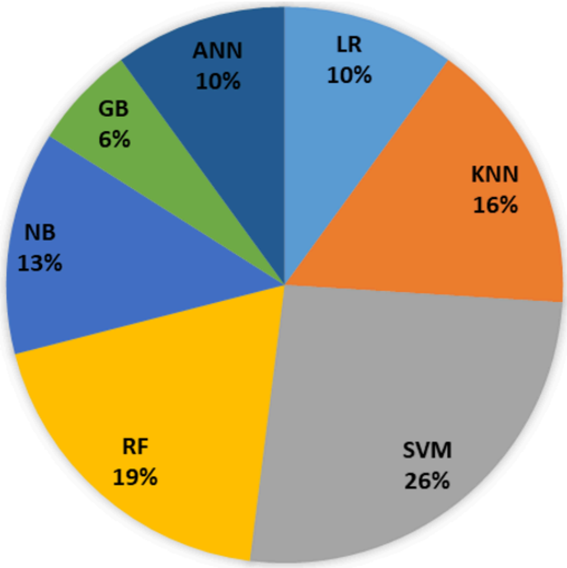


Fig. 4: Percentage of work done using different techniques on breast cancer disease

Figure (4) shows the percentage of work on breast cancer detection using various ML algorithms. It has been observed that SVM is the mostly used model for detecting breast cancer.

Table 3: Summarized review of breast cancer disease recommendation systems

Author	Dataset	Technique/Algorithm	Accuracy %
Allugunti (2022)	Kaggle	CNN	99.67
		SVM	89.84
		RF	90.55
Mridha (2021)	Wisconsin breast database from UCI	RF	98.83
		KNN	91.22
		ANN	99.73
		GB	96.49
		SVM	98.24
		LR	98.24
		NB	94.73
Islam *et al.* (2020)	Wisconsin breast cancer from UCI	SVM	97.1
		KNN	97.1
		ANN	98.57
		RF	95.71
		LR	95.71
Ak (2020)	University of Wisconsin Hospital	LR	98
		KNN	96
		SVM	97
		NB	95
		DT	95
		RF	95
		Rotation forest	96
Sharma *et al.* (2018)	Wisconsin diagnosis from UCI ML	RF	94.74
		KNN	95.90
		NB	94.47
Yarabarla *et al.* (2019)	WDBC(UCI)	KNN	70
		SVM	68
		RF	72
		GB	73
Omondiagbe *et al.* (2019)	WDBC from UCI MLR	SVM	96.47
		ANN	97.06
		NB	91.18
Tapak *et al.* (2019)	Breast cancer dataset from a retrospective cohort study	SVM and LDA	93
		AdaBoost	89
Tahmooresi *et al.* (2018)	Breast cancer dataset from mammogram images	SVM	99.8

Liver Disease

Choudhary *et al.* utilized the algorithms SVM, LR, RF and NB. The logistic classifier categorizes the data with a 71% overall accuracy by using a precise and accurate diagnosis of individuals at risk and early detection of liver diseases (Choudhary *et al.*, 2021). Singh *et al.* performed feature extraction and implemented a variety of machine learning techniques, including LR, KNN, XG-Boost, SVM, Gaussian NB, RF, DT, GB, CatBoost, AdaBoost and LightGBM. It was discovered that RF performed best among all the techniques, gained high accuracy, and performed outstandingly in all metric evaluations (Singh *et al.*, 2021). Rabbi *et al.* used ILPD dataset to apply four machine learning algorithms for categorizing liver patients. The Pearson Correlation Coefficient (PCC-FS) is utilized to refine the method and improve performance. On the algorithms LR, DT, RF and ET, Adaboost was applied. The performance of those algorithms is improved overall when Adaboost is used with PCC-FS. According to the study of experimental data, ET, the Adaboost algorithm's basic estimator, has the highest accuracy at 92.19% (Rabbi *et al.*, 2020). Hartatik *et al.* implemented NB and KNN algorithms to forecast liver disease. The (UCI) MLR was used to get the information. The maximum accuracy, 72.5%, was attained by the NB (Hartatik *et al.*, 2020). Devikanniga *et al.* used CSA-SVM classifier allowed for the accurate diagnosis of liver disease. It is observed that using CSA makes optimizing SVM easier than using SMO. The highest number, 99.49%, is produced by the CSA-overall SVM's classification accuracy (Devikanniga *et al.*, 2018). Wu *et al.* used RF, NB, ANN and LR to predict fatty liver disease. The study involved 577 patients in all and 377 of them had fatty livers. The highest level of RF accuracy was 87.48% (Wu *et al.*, 2019). Thirunavukkarasu *et al.* (2018) used many classification techniques, including LR, KNN and SVM on ILPD dataset downloaded from the UCL Machine Learning repository. LR and KNN outperformed with accuracy of 73.97% (Thirunavukkarasu *et al.*, 2018). Muthuselvan *et al.* used the liver patient dataset and machine learning algorithms like NB, K-star, J48 and Random Tree. The accuracy of the NB, K-star, J48 and RF was 60.6, 67.2, J48, 71.2 and 74.2% respectively (Muthuselvan *et al.*, 2018). Summary of studies discussed on liver disease prediction using various algorithms is given in Table (4).

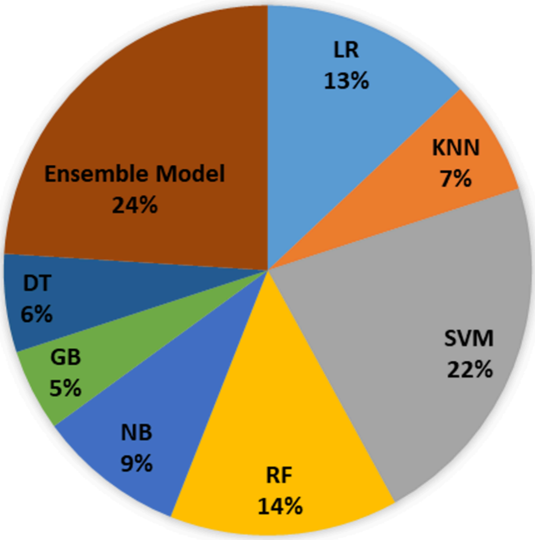


Fig. 5: Percentage of work done using different techniques on liver disease

Figure (5) shows the percentage of work on liver disease detection using various methodologies. It has been observed that the ensemble model is mostly used for the detection of liver disease.

Rehman *et al.* (2017) presented a cloud-based system to recommend diets based on users' pathological reports. It utilized the ant colony optimization algorithm to generate optimal food lists, aiming to fulfill nutritional requirements and manage diseases. The system demonstrated significant improvements in processing speed when executed in parallel on the cloud.

Table 4: Summarized review of liver disease recommendation systems

Author	Dataset	Technique/Algorithm	Accuracy %
Choudhary. (2021)	various medical laboratories, hospitals, and medical centers	SVM	68.49
		LR	70.54
		RF	69.86
		GB	67.88
		NB	58.90
Singh *et al.* (2021)	Kaggle	LR	72.89
		KNN	73.49
		XG-BOOST	86.75
		SVM	66.27
		GNB	71.69
		RF	88.55
		DT	83.13
		GB	80.12
		CATBOOST	81.93
		ADABOOST	72.89
		LIGHTGBM	84.34
Rabbi *et al.* (2020)	ILPD from the (UCI) MLR	LR	70.68
		DT	87.26
		RF	87.86
		ET	92.19
Hartatik *et al.* (2020)	ILPD from (UCI)MLR	NB	72.5
		KNN	63.19
Devikanniga *et al.* (2018)	ILPD from (UCI)MLR	CSA-SVM	99.49
		MVO-SVM	95.03
		GA-SVM	94.51
		FA-SVM	86.62
		PSO-SVM	85.25
Wu *et al.* (2019)	New Taipei City Municipal Hospital	RF	92.5
		NB	88.8
		ANN	89.5
		LR	85.4
Singh *et al.* (2018)	ILPD from (UCI)MLR	LR	73.97
		KNN	73.97
		SVM	71.97
Muthuselvan *et al.* (2018)	Andhra Pradesh's North East area	NB	60.6
		K-star	67.2
		J48	71.2
		RF	74.2

The author (Sookrah *et al.*, 2019) proposed a DASH diet recommender system that recommended healthy Mauritian diet plans to hypertensive patients. The system consisted of a recommendation engine that used techniques such as content-based filtering along with machine learning algorithms to recommend personalized diet plans based on factors such as age, user food preferences, allergies, smoking level, alcohol level, blood pressure level, and dietary intake. The system utilized a mobile application, which was handy and quick to use.

The authors (Bhat and Ansari, 2021) developed an Android-based healthcare app designed to quickly detect various diseases. They created a novel, hybrid machine-learning algorithm combining Support Vector Machine (SVM) and Genetic Algorithm (GA) to improve accuracy and reduce the complexity and number of database attributes. The algorithm's performance was analyzed using statistical parameters such as accuracy, confusion matrix, and ROC curve. The primary goal of this research was to predict diseases based on specific details. For complex diseases like cancer, the app enables quick detection, assisting doctors in initiating the appropriate treatment promptly. Additionally, the app recommends diets tailored to the patient's profile.

The proposed work focuses on two common diseases: Diabetes and Cardio Vascular Disease (CVD). A supervised classification algorithm is used to predict diseases, and an unsupervised clustering algorithm is employed to make diet recommendations. The objective is to enable convenient disease prediction at home and offer personalized, healthy diet recommendations (Lakshmi *et al*., 2024).

Methodology

In the proposed system IntelliHealth, data has been collected for four distinct diseases and ensemble approach has been suggested for detection diseases. To enhance user experience, a user-friendly web-based interface have been implemented. This interface not only facilitates disease detection but also offers personalized diet plans based on individual symptoms.

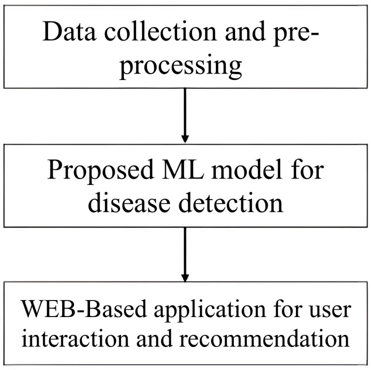


Fig. 6: Flow diagram of the proposed methodology

The proposed methodology of the proposed system, i.e., IntelliHealth for disease detection and recommendation, is given in Figure (6).

The proposed methodology utilizes machine learning classification algorithms and performance analysis is done to predict four distinct diseases. The flow diagram of the proposed system is given in Figure (7).

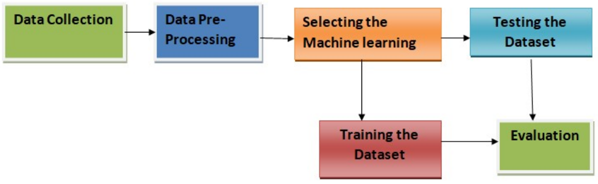


Fig. 7: Proposed workflow for proposed machine learning approach for disease detection

Data Collection

In this study, 4 diseases that is diabetes, breast cancer, thyroid and liver has been considered and dataset are collected from public repositories.

Diabetic Dataset

The diabetes dataset taken from the Pima Indian Diabetes Database (Pima Indians Diabetes Database). This dataset has 768 instances and 9 features. The features include glucose, BMI, pregnancy, skin thickness, insulin, age and blood pressure. The outcome variable predicts whether the patient is diabetic or not.

Breast Cancer Dataset

Wisconsin breast cancer diagnostic data is utilized to assess whether the cancer is benign or malignant (Breast Cancer Wisconsin (Diagnostic) Data Set). This dataset has 569 instances and 33 features.

Thyroid Dataset

The Garvan Institute in Sydney, Australia, provided information for the thyroid database (Thyroid Disease Data Set). This dataset has 3772 instances and 30 features. The features contain sex, antithyroid medication, sick, pregnant, thyroxine, thyroid surgery, T131 treatment, query hypothyroid, age, TT4 measured, T4U, RTI, TBG Measured, referral source and binary class.

Liver Disease Dataset

The disease dataset for the liver patient was obtained from the UCI ML repository (Indian Liver Patient Records). This dataset has 583 instances and 11 features. This dataset contains 416 records for liver patients and 167 records for non-liver patients.

Data Pre-Processing

Data pre-processing is required in order to reduce noise and build a more accurate and robust ML model. It is the process of cleaning up the data. The liver disease dataset has total 583 instances and 11 features, it contains 4 missing values. After removing the missing values, the final outcomes of this dataset have 579 instances. The thyroid dataset has total 3772 instances and 30 features, it contains 1129 missing values. The final values of this dataset have 2643 instances and 29 features, after removing the missing values. The breast cancer and diabetes dataset does not have any null values as shown in Table (5). So, the values of breast cancer dataset have 569 instances and 33 features. The diabetes dataset has 768 instances and 9 features. After removing the missing values from dataset, categorical data is numerical data. To apply machine learning algorithms, the dataset is split into training and testing sets. Table (5) provides the analysis of different datasets before handling missing values and after that. Feature scaling is performed on the processed datasets using standardization.

Table 5: Dataset size before and after removing missing values

Dataset	Actual size	Missing values	Dataset size after removing missing values
Liver disease dataset (Indian Liver Patient Records)	583 instances,11 features	4	579 instances, 11 features
Thyroid disease dataset (Thyroid Disease Data Set)	3772 instances 30 features	1129	2643 instances, 29 features
Breast cancer disease dataset (Breast Cancer Wisconsin (Diagnostic) Data Set)	569 instances,33 features	0	569 instances, 33 features
Diabetes disease dataset (Pima Indians Diabetes Database)	768 instances,9 features	0	768 instances, 9 features

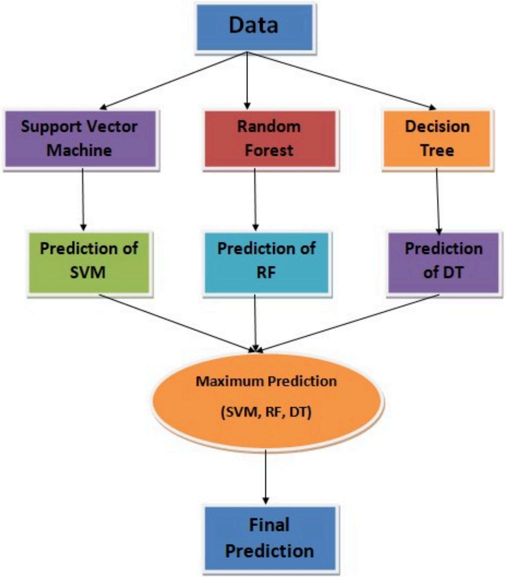


Fig. 8: Proposed ensemble model

Proposed Model

In this study, an ensemble-based technique has been proposed in which three ML classifiers that is 5 SVM, DT and RF have been combined. By combining predictions from many models, ensemble learning 6 is ML approach that improves predicting accuracy and robustness. Through the use of the ensemble's 7 collective intelligence, it seeks to reduce any inaccuracies or biases that may be present in individual 8 models. For classification issues, the max voting approach is often utilized. Figure (8) shows the proposed 9 ensemble model for predicting liver, thyroid, diabetes and breast cancer diseases.

Results and Discussion

To implement each of the models, we used the scikit-learn library, available for the Python language. The evaluation metrics and hyper-parameter tuning of diabetes, breast cancer, thyroid and liver are presented as follows:

Evaluation Metrics

An accuracy score is a measure of a model's performance. It is calculated by dividing the total of the true positive and true negative values by the total of the false positive and false negative values. The formula is given in Eq. (1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Hyper-Parameter Tuning

The ML models are evaluated via hyper-parameter tuning, which is the process of selecting a set of ideal hyper-parameters. The proposed system helped to tune the hyper-parameters using the grid search approach to the three mentioned classification algorithms.

Diabetes

In the diabetes experiment for SVM, the parameters C, gamma and Kernel are used. With the parameter C = 1, gamma = 0.0001 and kernel = "rbf", it achieved an accuracy of 79.5% much higher compared to the other parameters as shown in Table (6).

For RF classifier, the parameters Criterion and number of estimators are used. With the best parameter criterion = entropyandn-estimators =7 0, this technique achieves 80.3% accuracy.

For the Decision Tree classifier, the parameters criterion and spliter are used. With the best parameter, criterion = 'entropy' and spliter = 'best', it has achieved 75.4% accuracy.

The proposed ensemble model combined with the best parameter values of SVM, DT and RF classifier; it provides 100% accuracy.

Figure (9) presents the comparison of proposed ensemble their base estimators. Table (7) presents the comparative analysis of proposed method with existing techniques.

Table 6: Classification performance on diabetes of different classifier with respect to parameters

Predictive model	Hyper parameters			Accuracy %
SVM	C	gamma	Kernel	
	0.1	0.0001	rbf	76.4
	1	0.0001	rbf	79.5
	10	0.0001	rbf	78.7
	100	0.0001	rbf	78.9
RF	Criterion		n-estimators	
	Gini		10	76.4
	Gini		30	76.4
	Gini		50	77.9
	Gini		100	80.3
	Entropy		10	78.9
	Entropy		30	78.7
	Entropy		50	77.2
	Entropy		70	80.3
	Entropy		100	77
	Log_loss		10	77.9
	Log_loss		30	78.7
	Log_loss		90	78
	Log_loss		100	78.9
DT	Criterion		Splitter	
	Gini		Best	73
	Entropy		Best	75.4
	Entropy		Random	68.3
	Log_loss		Best	74.6
	Log_Loss		Random	71.5
Proposed ensemble	SVM (kernel = 'rbf', C = 1, gamma = 0.0001), DT (criterion = 'Entropy', Splitter = 'Best'), RF (criterion = 'Entropy', 'n_estimators' = 70)			100

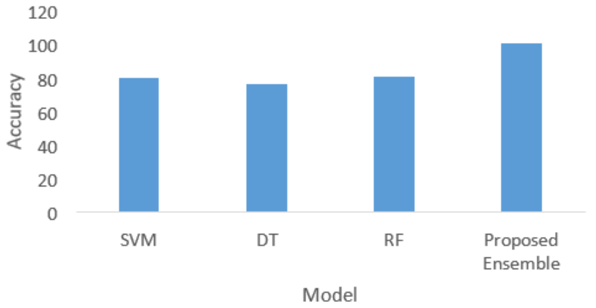


Fig. 9: Comparative analysis of proposed ensemble with SVM, DT and RF for diabetes disease detection

Table 7: The comparative analysis of existing study on the Pima Indian diabetes dataset with proposed method

Author/Techniques/References	Accuracy (%)
Krishnamoorthi *et al.* (2022)	91
Alaa Khaleel and Al-Bakry (2023)	94
Rajput *et al.* (2022)	96
Nadeem *et al.* (2021)	93.63
Choudhury and Gupta (2019)	77
Alehegn *et al.* (2018)	90.36
Dey *et al.* (2018)	82.35
Proposed ensemble	100

Breast Cancer

Using the parameters C, gamma and Kernel for the SVM, with C = 1000, gamma = 0.1 and kernel = rbf, it has achieved 97.4% accuracy and it performs better than other parameters show in Table (8).

Table 8: Classification performance on breast cancer of different classifier with respect to parameters

Predictive model	Hyper parameters			Accuracy %
SVM	C	Gamma	Kernel	
	0.1	0.1	rbf	96.1
	1	0.001	rbf	94.7
	10	0.01	rbf	96.1
	100	0.01	rbf	96.1
	1000	0.1	rbf	97.4
RF	Criterion		n-estimators	
	Gini		10	96.1
	Entropy		10	97.4
	Log-loss		10	96.1
DT	Criterion		Splitter	
	Gini		Best	96.1
	Gini		Random	94.8
	Entropy		Best	94.8
	Entropy		Random	96.1
	Log-loss		Best	96.1
	Log-Loss		Random	93.4
Ensemble	SVM (kernel = 'rbf', C = 1000, gamma = 0.1), DT (criterion = 'gini', Splitter = 'Best'), RF (criterion = 'Entropy', 'n-estimators' = 10)			99

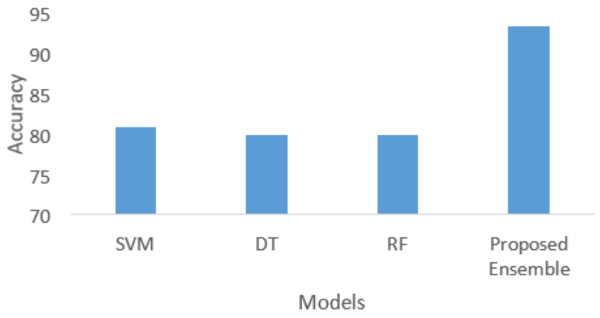


Fig. 10: Comparative analysis of proposed ensemble with its base estimators

Table 9: The comparison analysis of existing study on the Wisconsin breast cancer dataset

Author/Techniques/References	Accuracy
Islam *et al.* (2020)	98.57
Ak (2020)	98
Sharma *et al.* (2018)	95.90
Yarabarla *et al.* (2019)	73
Omondiagbe *et al.* (2019)	97.06
Proposed ensemble	99

Table 10: Performance of different classifier for Thyroid disease detection with respect to different hyper parameters

Predictive model	Hyper parameters			Accuracy %
SVM	C	Gamma	Kernel	
	0.1	1	rbf	97.6
	1	0.1	rbf	97.2
	10	0.1	rbf	97.2
	100	1	rbf	97.6
RF	Criterion		n-estimators	
	Gini		10	98.2
	Gini		30	98.2
	Entropy		10	97.6
	Entropy		30	97.6
	Log_loss		10	98.2
DT	Criterion		Splitter	
	Gini		best	96.5
	Entropy		Best	96.5
	Entropy		Random	95.4
	Log-loss		Best	96.7
Ensemble	SVM (kernel = 'rbf', C = 0.1, gamma = 1), DT (criterion = 'gini', Splitter = 'Best'), RF (criterion = 'gini', 'n-estimators' = 30)			100

For the RF classifier the parameters criterion and number of estimators, are applying. With the best parameter values criterion = 'entropy' and number of estimators = 10, it has provided 97.4% accuracy.

For the DT classifier the criterion and split parameter are used. With the best parameter values criterion = 'gini' and spliter = 'best', it has provided 96.1% accuracy.

The proposed ensemble model combined with best parameter values of SVM, RF and DT classifier; it provides accuracy of 90%. Figure (10) shows the comparison of proposed ensemble with other ML models for detection of breast cancer. Table (9) presents the comparison of the proposed method with already available techniques.

Thyroid

In the thyroid experiment for SVM, the parameters C, gamma and Kernel are used. With the parameter C = 1, gamma =0.1 and kernel = "rbf", it obtained100% accuracy for thyroid illness in Table (10).

For the RF classifier the parameters Criterion and number of estimators are used. With the best parameter values, criterion = 'gini' and number of estimators =30, it achieved an accuracy of 100%.

For the DT classifier, criterion and spliter parameter are used. With the best parameter values criterion ='gini' and spliter = 'best', it obtained an accuracy of 100%.

The proposed ensemble model combined with best parameter values of SVM, DT and RF classifier, it achieved an accuracy of 100%. Figure (11) shows the comparative analysis of proposed ensemble with DT, SVM and RF for thyroid detection. Comparative analysis of proposed ensemble model with already existing approaches is given in Table (11).

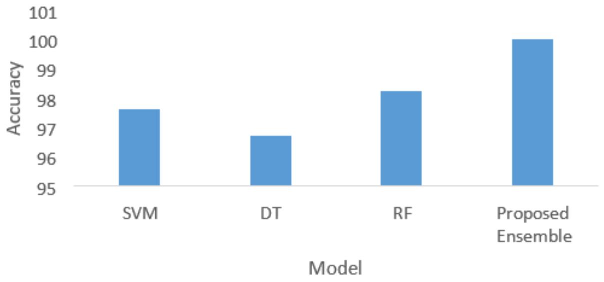


Fig. 11: Comparative analysis of proposed ensemble on thyroid disease detection with its base estimators

Table 11: Comparative analysis of proposed ensemble with existing study for detection of thyroid disease

Author/Techniques/References	Accuracy (%)
Shivastuti *et al.* (2021)	92
Chaubey *et al.* (2021)	96.87
Raghuraman *et al.* (2019)	97.35
Raisinghani *et al.* (2019)	99.46
Tyagi *et al.* (2018)	99.63
Shankar *et al.* (2020)	97.49
Proposed ensemble	100

Table 12: Comparative analysis of proposed ensemble with its base estimator for the detection of liver disease

Predictive model	Hyper parameters			Accuracy %
SVM	C	Gamma	Kernel	
	0.1	1	rbf	74.2
	1	0.1	rbf	75
	10	0.0001	rbf	74.2
	100	0.01	rbf	80.6
RF	Criterion		n-estimators	
	Gini		10	76.8
	Gini		30	78.3
	Gini		70	79.7
	Gini		90	73.9
	Entropy		10	78.3
	Entropy		70	75.4
	Entropy		100	76.8
	Log-loss		10	73.9
	Log_loss		50	78.3
	Log-loss		70	76.8
DT	Criterion		Splitter	
	Gini		best	73.9
	Gini		Random	72.5
	Entropy		Best	79.7
	Entropy		Random	78.3
	Log-loss		Best	76.8
	Log-loss		Random	66.7
Proposed ensemble	SVM (kernel = 'rbf', C = 100, gamma = 0.01), DT (criterion = 'Entropy', Splitter = 'Best'), RF (criterion = 'gini', 'n-estimators' = 70)			93

By combining the optimal SVM, DT and RF classifier parameter values, the proposed ensemble model was able to achieve 100% accuracy. The proposed ensemble model for thyroid detection is compared with DT, SVM and RF as presented in Figure (11). Table (11) provides a comparative analysis of the proposed ensemble model and existing approaches.

Liver

Using the parameters C, gamma and Kernel for the SVM, with c = 100, gamma = 0.01 and kernel = rbf, it achieved 80.6% of accuracy. It performs better than other parameters as shown in Table (12).

For the RF classifier the parameter criterion and number of estimators are used. With the best parameter values, criterion = 'gini' and no. of estimators=70, it has provided 79.7% accuracy.

For the DT classifier the criterion and spliter parameter are used. With the best parameter values criterion = 'entropy' and spliter = 'best', it has provided 79.7% accuracy.

The proposed ensemble model combined with best parameter values of DT, SVM and RF classifier; it provides the greatest accuracy of 93%. The comparative analysis of proposed ensembled model with other ML models is presented in Figure (12). A comparison on proposed ensemble with existing techniques is given in Table (13).

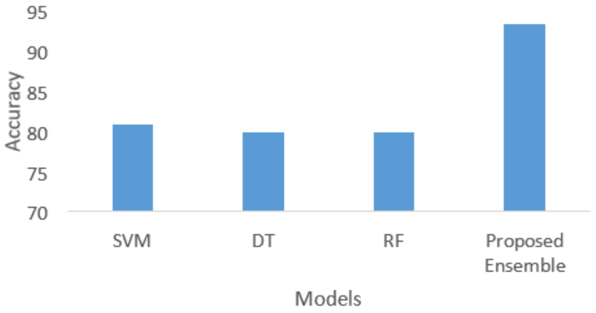


Fig. 12: Comparative analysis of proposed ensemble with its base estimator for the detection of liver disease

Table 13: The comparison analysis of existing work on the liver dataset from the (UCI) MLR

Author/Techniques/References	Accuracy (%)
Rabbi *et al.* (2020)	92.19
Hartatik *et al.* (2020)	72.5
Singh *et al.* (2018)	73.97
Proposed ensemble	93

Web Application for Various Disease Prediction

In recent years, numerous systems have been developed to integrate disease detection with dietary recommendations. However, IntelliHealth distinguishes itself through its innovative hybrid machine learning algorithm, which combines Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF). This approach not only improves prediction accuracy but also simplifies the dataset by reducing redundant features. Moreover, the IntelliHealth web-based application offers a seamless and instant user experience for disease detection and personalized diet planning. Figures (13-18) shows the working interfaces of proposed recommender system.

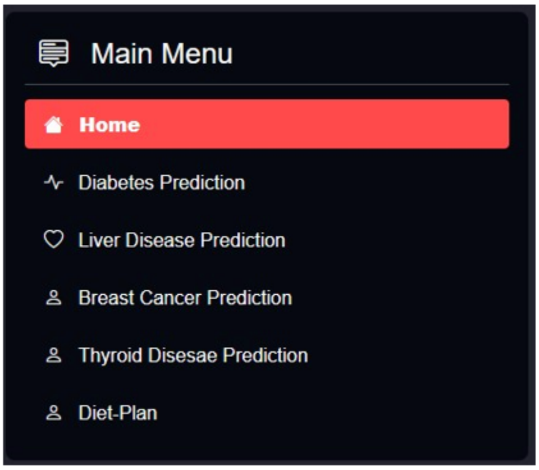
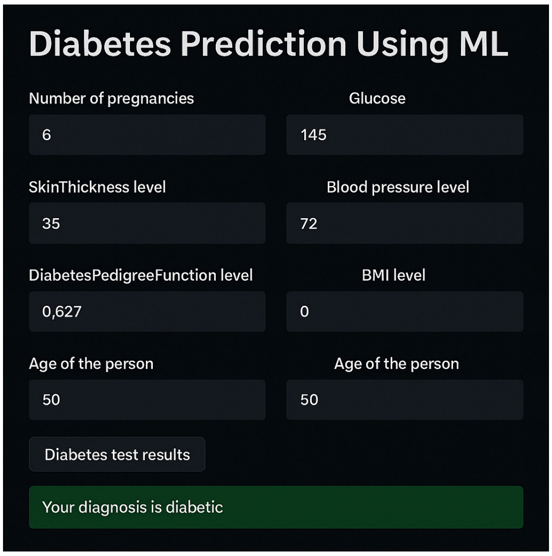


Fig. 13: Main Menu


An Intelligence based Healthcare Recommendation System Using Machine Learning

An intelligence based healthcare recommendation system predicts and recommends the various diseases accurately using machine learning models with the ensemble technique on healthcare datasets.This e-healthcare website will provide diet plans based on the patient's Profile.

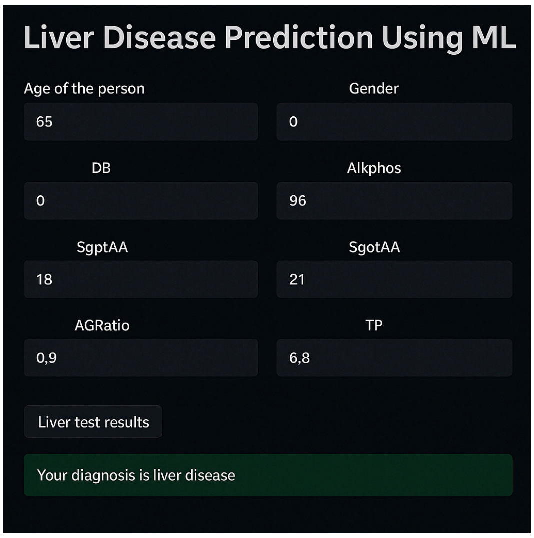
Fig. 14: Home page


Diabetes Prediction Using ML

Number of pregnancies: 6
Glucose: 145
SkinThickness level: 35
Blood pressure level: 72
DiabetesPedigreeFunction level: 0,627
BMI level: 0
Age of the person: 50
Age of the person: 50

Your diagnosis is diabetic

Fig. 15: Diabetes prediction


Liver Disease Prediction Using ML

Age of the person
65

Gender
0

DB
0

Alkphos
96

SgptAA
18

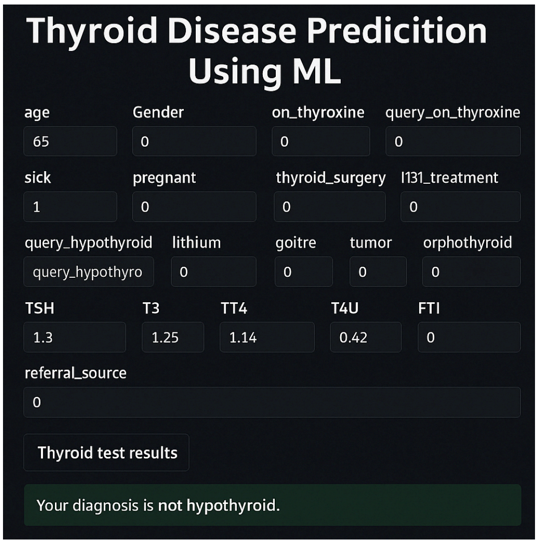
SgotAA
21

AGRatio
0,9

TP
6,8

Your diagnosis is liver disease

Fig. 16: Liver prediction


Thyroid Disease Predicition Using ML

age
65

Gender
0

on_thyroxine
0

query_on_thyroxine
0

sick
1

pregnant
0

thyroid_surgery
0

I131_treatment
0

query_hypothyroid
query_hypothyro

lithium
0

goitre
0

tumor
0

orphothyroid
0

TSH
1.3

T3
1.25

TT4
1.14

T4U
0.42

FTI
0

referral_source
0

Your diagnosis is not hypothyroid.

Fig. 17: Thyroid prediction

Fig. 18: Diet-plan for diseases

Conclusion

This study proposed ensemble-based machine learning strategies for predicting thyroid, breast cancer, liver and diabetic diseases. The datasets for this investigation were obtained from Kaggle. The performance of the proposed system was evaluated on the basis of accuracy. The results showed that the proposed ensemble model performed better than other ML based models. Additionally, a web-based application has been developed that uses patient information like age, gender and other factors to identify the precise ailment from which patients are suffering. The application suggests a diet plan and the type of exercise for the patient after diagnosing the disease.

Potential limitations include the reliance on specific datasets, which may not capture all demographic variations. Future research should explore additional datasets and refine the model for broader applicability. Additionally, developing a mobile application and incorporating deep learning techniques could further improve system performance.

Acknowledgment

The author, Sandeep Kaur, sincerely acknowledges the invaluable guidance and support of Dr. Ankita Wadhawan, Dr. Usha Mittal, and Dr. Priyanka Chawla, whose expertise and mentorship significantly contributed to the successful completion of this research.

Funding Information

The authors have not received any financial support or funding to report.

Ethics

The authors affirm that there are no conflicts of interest and that the work complies with the ethical policies of the Journal of Computer Science.

References

Abbad Ur Rehman, H., Lin, C.-Y., & Mushtaq, Z. (2021). Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease. *Journal of the Chinese Institute of Engineers*, *44*(1), 77-87. https://doi.org/10.1080/02533839.2020.1831967

Abbad Ur Rehman, H., Lin, C.-Y., Mushtaq, Z., & Su, S.-F. (2021). Performance Analysis of Machine Learning Algorithms for Thyroid Disease. *Arabian Journal for Science and Engineering*, *46*(10), 9437-9449. https://doi.org/10.1007/s13369-020-05206-x

Ak, M. F. (2020). A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. *Healthcare*, *8*(2), 111-187. https://doi.org/10.3390/healthcare8020111

Alaa Khaleel, F., & Al-Bakry, A. M. (2023). Diagnosis of Diabetes Using Machine Learning Algorithms. *Materials Today: Proceedings*, *80*, 3200-3203. https://doi.org/10.1016/j.matpr.2021.07.196

Alehegn, M., Joshi, R., & Mulay, P. (2018). Analysis and Prediction of Diabetes Mellitususing ML Algorithm. *International Journal of Pure and Applied Mathematics*, *118*(9), 871-878.

Allugunti, V. R. (2022). Breast Cancer Detection Based on Thermographic Images Using Machine Learning and Deep Learning Algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56. https://doi.org/10.33545/26633582.2022.v4.i1a.68

Bhat, S. S., & Ansari, G. A. (2021). Predictions of Diabetes and Diet Recommendation System for Diabetic Patients using Machine Learning Techniques. *2021 2nd International Conference for Emerging Technology (INCET)*, 1-5. https://doi.org/10.1109/incet51464.2021.9456365

Breast Cancer Wisconsin (Diagnostic) Data Set. (n.d.). https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

Chaubey, G., Bisen, D., Arjaria, S., & Yadav, V. (2021). Thyroid Disease Prediction Using Machine Learning Approaches. *National Academy Science Letters*, *44*(3), 233-238. https://doi.org/10.1007/s40009-020-00979-z

Choudhary, R., Gopalakrishnan, T., Ruby, D., Gayathri, A., Murthy, V. S., & Shekhar, R. (2021). An Efficient Model for Predicting Liver Disease Using Machine Learning. *Data Analytics in Bioinformatics: A Machine Learning Perspective*, 443-457.

Choudhury, A., & Gupta, D. (2019). A Survey on Medical Diagnosis of Diabetes Using MLTechniques. *Recent Developments in ML and Data Analytics*, *740*, 67-78. https://doi.org/10.1007/978-981-13-1280-9_6

Devikanniga, D., Ramu, A., & Haldorai, A. (2018). Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm. *EAI Endorsed Transactions on Energy Web*, *7*(29), 164177. https://doi.org/10.4108/eai.13-7-2018.164177

Dey, S. K., Hossain, A., & Rahman, Md. M. (2018). Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm. *2018 21st International Conference of Computer and Information Technology (ICCIT)*, 1-5. https://doi.org/10.1109/iccitechn.2018.8631968

Hartatik, H., Tamam, M. B., & Setyanto, A. (2020). Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms. *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, 1-5. https://doi.org/10.1109/icoris50180.2020.9320797

Indian Liver Patient Records. (n.d.). https://www.kaggle.com/datasets/uciml/indian-liver-patient-records

Islam, Md. M., Haque, Md. R., Iqbal, H., Hasan, Md. M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Computer Science*, *1*(5), 1-14. https://doi.org/10.1007/s42979-020-00305-w

Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *Journal of Healthcare Engineering*, *2022*(1), 1-10. https://doi.org/10.1155/2022/1684017

Lakshmi, K., Deeba, K., Harave, V., & Bharti, S. (2024). Life Expectancy Prediction And Diet Recommendation System for Cardiovascular and Diabetes Disease Using Machine Learning. *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, 1-8. https://doi.org/10.1109/ickecs61492.2024.10616598

Lella, L., Licata, I., & Pristipino, C. (2022). Pima Indians Diabetes Database Processing through EBBM-Optimized UTM Model. *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies*. 15th International Conference on Health Informatics, Online Streaming, --- Select a Country ---. https://doi.org/10.5220/0010806500003123

Mridha, K. (2021). Early Prediction of Breast Cancer by using Artificial Neural Network and Machine Learning Techniques. *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 582-587. https://doi.org/10.1109/csnt51715.2021.9509658

Muthuselvan, S., Rajapraksh, S., Somasundaram, K., & Karthik, K. (2018). Classification of Liver Patient Dataset Using Machine Learning Algorithms. *International Journal of Engineering And Technology*, *7*(3.34), 323-326. https://doi.org/10.14419/ijet.v7i3.34.19217

Nadeem, M. W., Goh, H. G., Ponnusamy, V., Andonovic, I., Khan, M. A., & Hussain, M. (2021). A Fusion-Based Machine Learning Approach for the Prediction of the Onset of Diabetes. *Healthcare*, *9*(10), 1393-1401. https://doi.org/10.3390/healthcare9101393

Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine Learning Classification Techniques for Breast Cancer Diagnosis. *IOP Conference Series: Materials Science and Engineering*, *495*, 012033. https://doi.org/10.1088/1757-899x/495/1/012033

Rabbi, Md. F., Mahedy Hasan, S. M., Champa, A. I., AsifZaman, Md., & Hasan, Md. K. (2020). Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study. *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, 111-116. https://doi.org/10.1109/icaict51780.2020.9333528

Raghuraman, M. T., Sailatha, E., & Gunasekaran, S. (2019). Efficient Thyroid Disease Prediction andComparative Study Using ML Algorithms. *International Journal of Information and Computing Science6*, *6*(6), 617-624.

Raisinghani, S., Shamdasani, R., Motwani, M., Bahreja, A., & Raghavan Nair Lalitha, P. (2019). Thyroid Prediction Using Machine Learning Techniques. *Advances in Computing and Data Sciences*, *1045*, 140-150. https://doi.org/10.1007/978-981-13-9939-8_13

Rajput, D. S., Basha, S. M., Xin, Q., Gadekallu, T. R., Kaluri, R., Lakshmanna, K., & Maddikunta, P. K. R. (2022). Providing Diagnosis on Diabetes Using Cloud Computing Environment to the People Living in Rural Areas of India. *Journal of Ambient Intelligence and Humanized Computing*, *13*(5), 2829-2840. https://doi.org/10.1007/s12652-021-03154-4

Ramesh, J., Aburukba, R., & Sagahyroon, A. (2021). A Remote Healthcare Monitoring Framework for Diabetes Prediction Using Machine Learning. *Healthcare Technology Letters*, *8*(3), 45-57. https://doi.org/10.1049/htl2.12010

Rehman, F., Khalid, O., Haq, N. ul, Khan, A. ur R., Bilal, K., & Madani, S. A. (2017). Diet-Right: A Smart Food Recommendation System. *KSII Transactions on Internet and Information Systems*, *11*(6), 2910-2925. https://doi.org/10.3837/tiis.2017.06.006

Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018). Prediction of Diabetes Using Machine Learning Algorithms in Healthcare. *2018 24th International Conference on Automation and Computing (ICAC)*, 1-6. https://doi.org/10.23919/iconac.2018.8748992

Shankar, K., Lakshmanaprabu, S. K., Gupta, D., Maseleno, A., & de Albuquerque, V. H. C. (2020). Retracted Article: Optimal Feature-Based Multi-Kernel SVM Approach for Thyroid Disease Classification. *The Journal of Supercomputing*, *76*(2), 1128-1143. https://doi.org/10.1007/s11227-018-2469-4

Sharma, S., Aggarwal, A., & Choudhury, T. (2018). Breast Cancer Detection Using Machine Learning Algorithms. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 114-118. https://doi.org/10.1109/ctems.2018.8769187

Shivastuti, Kour, H., Manhas, J., & Sharma, V. (2021). Performance Evaluation of SVM and Random Forest for the Diagnosis of Thyroid Disorder. *International Journal for Research in Applied Science and Engineering Technology*, *9*(5), 945-947. https://doi.org/10.22214/ijraset.2021.34381

Singh, V., Gourisaria, M. K., & Das, H. (2021). Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1-7. https://doi.org/10.1109/gucon50781.2021.9573803

Sookrah, R., Dhowtal, J. D., & Devi Nagowah, S. (2019). A Dash Diet Recommendation System for Hypertensive Patients Using Machine Learning. *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 1-6. https://doi.org/10.1109/icoict.2019.8835323

Sultana, S., Momen, A. A., Haque, M., Khandaker, M. H., & Sakib, N. (2021). An Approach for Developing Diabetes Prediction and Recommendation System. *International Journal of Computer Applications*, *174*(14), 20-28. https://doi.org/10.5120/ijca2021921033

Tahmooresi, M., Afshar, A., Rad, B. B., Nowshath, K. B., & Bamiah, M. A. (2018). Early Detection of Breast Cancer Using ML Techniques. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC*, *10*(3-2), 21-27.

Tapak, L., Shirmohammadi-Khorram, N., Amini, P., Alafchi, B., Hamidi, O., & Poorolajal, J. (2019). Prediction of Survival and Metastasis in Breast Cancer Patients Using Machine Learning Classifiers. *Clinical Epidemiology and Global Health*, *7*(3), 293-299. https://doi.org/10.1016/j.cegh.2018.10.003

Thirunavukkarasu, K., Singh, A. S., Irfan, M., & Chowdhury, A. (2018). Prediction of Liver Disease using Classification Algorithms. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 1-3. https://doi.org/10.1109/ccaa.2018.8777655

Thyroid Disease Data Set. https://www.kaggle.com/datasets/yasserhessein/thyroid-disease-data-set

Tyagi, A., Mehra, R., & Saxena, A. (2018). Interactive Thyroid Disease Prediction System Using Machine Learning Technique. *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 689-693. https://doi.org/10.1109/pdgc.2018.8745910

Vasile, C. M., Udriștoiu, A. L., Ghenea, A. E., Popescu, M., Gheonea, C., Niculescu, C. E., Ungureanu, A. M., Udriștoiu, Ș., Drocaş, A. I., Gruionu, L. G., Gruionu, G., Iacob, A. V., & Alexandru, D. O. (2021). Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods. *Medicina*, *57*(4), 395-398. https://doi.org/10.3390/medicina57040395

Wu, C.C., Yeh, W.C., Hsu, W.D., Islam, Md. M., Nguyen, P. A., Poly, T. N., Wang, Y.C., Yang, H.C., & Li, Y.C. (2019). Prediction of Fatty Liver Disease Using Machine Learning Algorithms. *Computer Methods and Programs in Biomedicine*, *170*, 23-29. https://doi.org/10.1016/j.cmpb.2018.12.032

Yarabarla, M. S., Ravi, L. K., & Sivasangari, A. (2019). Breast Cancer Prediction via Machine Learning. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 121-124. https://doi.org/10.1109/icoei.2019.8862533