Date available online: 13-02-2025
Vol. 2, Issue 2 (June December 2024)
This is an open-access article.

Health Predictions Redefined: The Impact of AI on Future Disease Diagnosis

Iqra Muneer[1,2*], Sadia Tariq[1], Muhammad Kashif[2]

[1]Department of Computer Science & Engineering, University of Engineering Technology Lahore, Narowal Campus, Lahore, Pakistan.
[2]Comsats University Islamabad, Sahiwal Campus, Sahiwal, Pakistan.

[*]Corresponding author: Iqra Muneer (e-mail: iqramuneer@uet.edu.pk)

Abstract— Healthcare professionals often apply a one-size-fits-all approach in patient care, potentially leading to misdiagnosis, suboptimal treatments, and higher healthcare costs. Machine-learning models have garnered attention for their ability to improve diagnostic accuracy, with numerous studies focusing on machine learning applications for individual disease predictions, such as Type II diabetes, heart disease, kidney disease, and hypertension. However, limited research has tackled the combined prediction of Type I diabetes (standard cases), Type II diabetes (gestational diabetes), and cardiovascular disease, presenting a significant research gap.

To address this gap, we introduce a set of benchmark corpora based on authentic patient records, targeting specific disease categories. The first contribution is a heart disease corpus containing 606 instances. The second and third contributions consist of two separate corpora, each with 849 instances: one focused on standard diabetes cases and the other on gestational diabetes cases. We evaluated these corpora with ten machine-learning algorithms and five deep-learning algorithms, rigorously comparing their performance across common metrics, including accuracy, precision, recall, and $F1$-score. Our results revealed high performance across all models, with top $F1$-scores of 0.785 using Random Forest, 0.790 with Gradient Boosting, and 0.994 using BiLSTM for the combined disease prediction. These findings suggest that the proposed datasets and models provide a robust foundation for accurate and scalable high-risk disease prediction, contributing a valuable, multidimensional approach to personalized patient care.The novelty of our approach lies in the creation and use of region-specific datasets for combined prediction of Type I diabetes, Type II diabetes (gestational), and cardiovascular disease, which has been minimally explored in existing research.

Index Terms—Heart disease prediction Diabetes prediction, Diabetes during Pregnancy, Cardiovascular disease, Healthcare management

1 Introduction

The prevalence of diabetes is on the rise worldwide, even in developed nations mainly due to obesity and stress related to modern lifestyle. Due to its widespread impact, combating diabetes requires a collective effort from healthcare providers, patients, families, and society. The disease incurs significant social, health, and economic costs [1]. The chronic condition of diabetes arises when the body is unable to either produce sufficient insulin or effectively use the insulin produced, leading to high levels of sugar in the blood [2, 3]. The root cause of this condition, also known as "x syndrome," is still not completely comprehended by medical professionals. Treatment of diabetes has traditionally focused on symptom monitoring rather than targeting the underlying cause. According to the World Health Organization, approximately 5 percent of the world's population is affected by diabetes, and this figure is on the rise. In developed countries, diabetes is most common among individuals over 65 years old, while in developing countries, the highest incidence occurs among those aged 45-64 years, with type II diabetes becoming more common among people aged 30-40 years [1]. To leverage the abundance of historical data, data mining techniques can be employed to detect patterns and trends in diabetes, facilitating early detection and prevention. By utilizing data mining, healthcare professionals can efficiently analyse pre-existing data to identify patterns and trends in diabetes. Healthcare providers and public health organizations can utilize the system to prevent and treat high-risk conditions such as diabetes, hypertension, and heart disease [15]. Healthcare practitioners to identify high-risk groups for chronic diseases and to devise tailored treatments to prevent and manage these disorders may use the method. The method may be used by public health organizations to analyse big datasets and identify risk factors for chronic diseases, which can aid in the creation of treatments and policies to prevent and manage these diseases [16, 17, and 18]. Cardiovascular illnesses include heart failure, coronary artery disease, stroke, and other conditions that affect the heart and blood vessels. Numerous factors, including high blood pressure, high cholesterol, smoking, obesity, diabetes, and a family history of the disease, can contribute to cardiovascular disease [13]. In other words, diabetes is a chronic illness that impairs the body's capacity to metabolize glucose or blood sugar. Type I diabetes, an autoimmune illness commonly identified in infancy, and type II diabetes, are linked to lifestyle related factors, including physical inactivity, obesity, and bad eating practices. Age, family history and poor nutrition are also risk factors for diabetes [14].

The disease prediction component of the system analyses a vast dataset of health and lifestyle parameters to estimate an individual's risk of acquiring heart disease andndiabetes using machine-learning algorithms. To give risk estimates diseases, these algorithms consider characteristics such as age, gender, BMI, blood pressure, cholesterol levels, smoking history, family medical history, and other pertinent data.

Several studies have investigated the use of machine learning for disease prediction type II diabetes prediction, Heart disease prediction, kidney disease prediction, and hypertension detection. However, for the combined prediction of type II (adult diabetes), and type III (Diabetes during Pregnancy) along with cardiovascular disease, formal study has been hardly carried out.

As a first major contribution, we develop a novel benchmark corpus based on real cases of heart patient records containing 606 instances. As another contribution, we have presented two novel benchmark corpora containing 849 instances based on real cases of diabetes in normal patients, and Diabetes during Pregnancy respectively. As a second contribution, all of these corpora were evaluated using 10 various machine learning algorithms Random Forest (RF), Decision Tree (DT), Bernoulli Naive Bayes (BNB), Gaussian Naive Bayes (GNB), Gradient Boosting Classifier (G-BC), AdaBoost (AB), Multilayer Perceptron (M-LP), K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), and Logistic Regression (LR). As another contribution, five different deep learning methods including Long Short Term Memory (LSTM), Bidirectional Long Short Term Memory (BILSTM), Convolution Neural Network (CNN), Gated Recurrent Unit (GRU), and Bidirectional Gated Recurrent Unit (BIGRU). As final and most fruitful contribution, an in-depth, and detailed comparison was performed among the applied algorithms. These datasets have been evaluated using Accuracy, Precision, Recall, and F_1-score. Overall, the proposed system will provide patients with Personalized healthcare management, Disease prediction, Diet recommendation, Performance evaluation and Feedback mechanism in a cost-effective manner. The research paper is organized as follows: Section 2 describes the literature review Section 3 covers the dataset creation methodology. Section 4 presents the experimental setup. Section 5 discusses and analyses the results, and Section 6 concludes the paper.

2 Literature Review

In literature, various efforts have been made to develop novel approaches and datasets for the task of various disease predictions. The following section contains the detail of some prominent attempts made for various disease prediction tasks.

Lahla et al. proposed a novel dataset for the prediction task of diabetes [4]. The dataset consists of 270 records from the Public Health Institute with seven different attributes including: 1) Age, 2) Body Mass Index, 3) Insulin, 4) Serum Insulin in two hours, 5) Glucose: Glucose tolerance test values, 6) Skin Thickness, 7) Blood Pressure, and 8) Number of pregnancies. The dataset was evaluated using machine learning techniques including Na¨ıve Bayes, Support Vector Machine, Decision Trees (DT), and Artificial Neural Networks (ANN). The best performance was obtained with an accuracy of 79% using DT.

Singh et al. proposed a study on heart disease prediction tasks [5]. The authors applied various approaches to a dataset from the UCI repository consisting of 304 records with 13 different attributes including: 1) Age, 2) Sex, 3) Chest pain, 4) Blood Pressure, 5) Fasting blood sugar, 6) Cholesterol, 7) Maximum electric cardiograph, 8) Heart rate, 9) Exercise angina, 10) Depression, 11) Slope of peak exercise segment, 12) Fluoroscopy, and 13) Defect type. The dataset was evaluated using machine learning methods including Logistic Regression (LR), K-Nearest Neighbors (KNN), and Random Forest (RF) Classifier. The best performance was obtained with an accuracy of 88.5% using KNN.

Another study [6] predicts chronic kidney disease using a dataset from UCI that contains 25 different characteristics. The dataset was analyzed using machine learning methods proposed by Pal and colleagues, including LR, DT, and Support Vector Machine (SVM). An accuracy of 97% was the best result obtained using DT.

Lukmanto et al. [7] conducted a study to forecast the onset of diabetes mellitus (DM) using 768 patient data points from the Pima Indian Diabetes Dataset. They utilized fuzzy support vector machines and feature selection to discover DM. Feature selection was used to locate relevant properties in the dataset, which was trained using SVM to provide fuzzy rules. The results showed an optimistic accuracy of 89

Mujumdar et al. proposed a novel dataset for diabetes prediction [8]. The dataset consists of 800 records with nine different input attributes including: 1) Age, 2) Body Mass Index, 3) Insulin, 4) Glucose, 5) Skin Thickness, 6) Blood Pressure, 7) Number of pregnancies, 8) Job type, and 9) Office work. Machine learning techniques such as DT, Gaussian Na¨ıve Bayes (GNB), Linear Discriminant Analysis (LDA), SVC, RF, Extra Trees, Ada Boost (AB), Multi-layer Perceptron (M-LP), LR, Gradient Boosting Classifier (G-BC), and KNN were used to analyze the dataset. The highest result was attained with an accuracy of 96% using Logistic Regression.

The task has been explored in multiple ways, including chronic kidney disease [6], diabetic mellitus [7], diabetes prediction [8, 4], and heart disease [5] with various algorithms including LR, KNN, RF, AB, and M-LP. However, the tasks have never been explored with a variety of deep learning methods. Furthermore, the task has not been explored on datasets based on Pakistan's national disease due to unavailability of datasets. To fulfill this gap, the study proposes a novel benchmark corpus comprising authentic heart patient records, totaling 606 instances. Additionally, we introduced two new benchmark corpora consisting of 849 instances each, derived from real cases of diabetes in both normal patients and during pregnancy. Another significant contribution lies in the evaluation of these corpora using a diverse set of machine learning algorithms, including RF, DT, BNB, GNB, G-BC, AB, M-LP, K-NN, SVM, and LR. Furthermore, we explored five distinct deep learning methodologies, namely LSTM, BILSTM, CNN, GRU, and BIGRU. Finally, our most substantial contribution involves an exhaustive and detailed comparative analysis of the performance of these applied algorithms.

Table 1 Comparison of Disease Prediction Studies

Study	Disease Type	Dataset	Algorithms Used	Key Findings / Novelty
Llaha & Rista [4]	Diabetes Prediction (Type II)	270 records from Public Health Institute, 7 attributes	Naïve Bayes, SVM, DT, ANN	Achieved 79% accuracy with DT; single disease focus
Singh & Kumar [5]	Heart Disease Prediction	304 records, 13 attributes from UCI dataset	LR, KNN, RF	Achieved 88.5% accuracy with KNN; heart disease only, no focus on other diseases
Pal [6]	Chronic Kidney Disease Prediction	UCI dataset with 25 attributes	LR, DT, SVM	Achieved 97% accuracy with DT; specific to kidney disease
Lukmanto et al. [7]	Diabetes Mellitus Prediction	Pima Indian dataset, 768 records	Fuzzy SVM, Feature Selection	Achieved 89% accuracy; focused on diabetes mellitus only
This Study	Combined prediction of Type I & Type II Diabetes (normal & gestational) with Cardiovascular Disease	Three real-world corpora: Heart disease (606 instances), Normal diabetes (849), Diabetes during pregnancy (849)	ML algorithms: RF, DT, SVM, GBC, KNN, etc.; DL models: LSTM, BiLSTM, CNN, GRU, BiGRU	**F_1-score: RF = 0.785, GBC = 0.790, BiLSTM = 0.994**; Unique in combining multiple disease types, region-specific datasets, benchmark corpora for diabetes/heart disease, and extensive ML/DL comparative analysis

In contrast to prior studies that primarily focus on predicting a single disease, such as diabetes or heart disease, this study uniquely combines Type I and Type II diabetes—including gestational diabetes—with cardiovascular disease prediction in a unified framework, addressing a significant gap in the literature (Table 1). By leveraging real-world, region-specific patient data from Pakistan, the study introduces a novel dataset that is tailored to an underrepresented population, enhancing its relevance and applicability. Notably, three benchmark corpora were developed as part of this research: one for heart disease cases (606 instances), another for normal diabetes cases (849 instances), and a third specifically for diabetes during pregnancy (849 instances). Furthermore, the study undertakes an extensive evaluation of traditional machine learning (ML) and deep learning (DL) models, employing a total of ten ML and five DL algorithms. This comprehensive approach provides nuanced insights into the most effective methods for each disease category, offering a valuable comparative performance analysis. The study achieved an impressive F_1-score of 0.994 with BiLSTM in combined prediction tasks, highlighting the superior predictive power of deep learning techniques over previous studies. Together, these findings underscore the study's contributions to the field by presenting an innovative, multidimensional dataset and a robust algorithmic framework that holds promise for more accurate and scalable high-risk disease prediction.

A	B	C	D	E	F	G	H	I
Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1

Figure 1 Diabetes (During Pregnancy) Dataset

3 Dataset Creation Methodology

This section sheds light that how all proposed corpora were created.

3.1 Diabetes (During Pregnancy) Dataset Collection

In this respect, we gathered information from several hospitals in Lahore, Pakistan, on diabetes in pregnant women. The information on numerous demographic, clinical, and lifestyle aspects was gathered in the form of CSV files. Gestational diabetes, commonly referred to as diabetes during pregnancy, is a disease when a woman experiences high blood sugar levels while she is pregnant. After birth, the disease often goes away. It commonly develops in the second or third trimester. Diabetes in pregnant women dataset contains 849 instances with 9 attributes, including age, number of times pregnant, glucose concentration, blood pressure, skin thickness, insulin level, body mass index, diabetes pedigree function, and the presence or absence of diabetes.

1. Age: This attribute indicates the age of the pregnant mother in years. It has a numerical quality.
2. Pregnancy frequency: This element indicates the pregnant woman's frequency of pregnancies. It has a numerical quality.
3. Glucose concentration: This factor shows the

UCP JOURNAL OF
ENGINEERING AND TECHNOLOGY

milligrams per decilitre (mg/dL) glucose level in the blood plasma of the expectant mother. It has a numerical quality.

4. Blood pressure: This feature displays the pregnant woman's diastolic blood pressure (in mm Hg). It has a numerical quality.

5. Skin thickness: The thickness of the skin on the triceps of a pregnant woman is measured (in millimetres). It has a numerical quality.

6. Insulin level: This factor reveals the amount of insulin (measured in " U/ml) in a pregnant woman's blood. It has a numerical quality.

7. Body mass index (BMI): This measurement reveals that the body mass index of expecting mother is calculated by dividing her weight in kilograms by her height in meters squared. It has a numerical quality.

8. This property is a representation of the diabetes pedigree function, which estimates the likelihood of diabetes based on family history.

9. Diabetic status: This trait reveals whether or not the expectant mother has diabetes. It is a binary attribute, where a value of one indicates the presence of diabetes and a value of zero indicates the lack of it.

	A	B	C	D	E	F	G	H
1	Gender	Age	BloodPres	SkinThick	Insulin	BMI	Glucose	Outcome
2	1	50	72	35	0	33.6	148	1
3	1	31	66	29	0	26.6	85	0
4	2	32	64	0	0	23.3	183	1
5	2	21	66	23	94	28.1	89	0
6	2	33	40	35	168	43.1	137	1
7	1	30	74	0	0	25.6	116	0
8	1	26	50	32	88	31	78	1
9	2	29	0	0	0	35.3	115	0
10	1	53	70	45	543	30.5	197	1
11	1	54	96	0	0	0	125	1
12	1	30	92	0	0	37.6	110	0
13	2	34	74	0	0	38	168	1
14	1	57	80	0	0	27.1	139	0
15	1	59	60	23	846	30.1	189	1
16	2	51	72	19	175	25.8	166	1
17	1	32	0	0	0	30	100	1
18	2	31	84	47	230	45.8	118	1
19	2	31	74	0	0	29.6	107	1
20	2	33	30	38	83	43.3	103	0

diabetesN

Figure 2 Normal Diabetes Dataset

3.2 Normal Diabetes Dataset Collection

In this dataset, information from several hospitals was collected in Lahore, Pakistan, from regular diabetics. The information on numerous demographic, clinical, and lifestyle aspects was gathered in the form of CSV files. The normal diabetes dataset contains 849 tuples with 8 attributes, including age, gender, glucose concentration, blood pressure, skin thickness, insulin level, body mass index, and the presence or absence of diabetes.

3.3 Heart Disease Dataset Collection

In this respect, information from several hospitals was collected in Lahore, Pakistan, from heart patients. The information on numerous demographic, clinical, and lifestyle aspects was gathered in the form of CSV files. The heart disease dataset has 606 tuples with 14 attributes in it, including age, sex, the type of chest pain, resting blood pressure, cholesterol level, fasting blood sugar, electrocardiogram results, maximum heart rate reached, exercise-induced angina, ST depression caused by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels coloured by fluoroscope, thallium stress test results, and absence or presence of heart disease.

3.3.1 Dataset Standrization

All of above proposed datasets are standardized in CSV format and is readily available for research purposes. It is licensed under a Creative Commons CC-BY-NC-SA license, allowing for free and open use while ensuring proper attribution and prohibiting commercial use without permission. This corpus can be accessed from the available link for the reviewers.

1. Age
2. Sex
3. Type of chest pain experienced (categorized into 4 values)
4. Resting blood pressure measurement
5. Serum cholesterol level in mg/dl
6. Fasting blood sugar level, with values greater than 120 mg/dl indicating the presence of high blood sugar
7. Results of a resting electrocardiogram (ECG), with values categorized as 0, 1, or 2
8. Maximum heart rate achieved during exercise
9. Presence or absence of exercise-induced angina
10. ST depression induced by exercise relative to rest (known as old peak)
11. Slope of the peak exercise ST segment
12. Number of major blood vessels (ranging from 0-3) colored by fluoroscopy.
13. thal: 0 = normal; 1 = fixed defect; 2 = reversible defect

Figure 3 Normal Diabetes Dataset Parameters

A	B	C	D	E	F	G	H	I	J	K	L	M	N
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
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1

Figure 4 Heart Disease Dataset

4 Experimental Setup

This section presents the experimental setup, including applied algorithms, evaluation measures, and evaluation methodology for high disease prediction tasks including diabetes predictions, and heart disease prediction.

4.1 Evaluation Measures

The most commonly used evaluation measures include: Accuracy, Precision, Recall, and F_1 measures. Accuracy is calculated by dividing the number of accurate predictions by the total number of predictions, the model produced [9]. The formula for Precision is given below.

$A = (TP + FP)/(TP + FP + FN + TN)$ (1)

Precision (P) can be defined as the proportion of true positive predictions from all the positive cases [10].

$p = TP/(TP + FP)$ (2)

Recall (R) is defined as the proportion of correctly identified positive cases [11].

$R = TP/(TP + FN)$ (3)

F_1 measure is the harmonic mean of precision (P) and recall (R). F_1 measure is commonly used as an evaluation measure for cases where datasets are unbalanced. It is defined as the harmonic mean of two other measures, Precision (P) and Recall (R) [12].

$F_1 = (2 * P * R)/(P + R)$ (4)

4.2 Evaluation Methodology

For both normal diabetes and diabetes during pregnancy, the problem with diabetes prediction was handled as a supervised text classification task. Two degrees of discrimination were intended by the categorization task: (1) diabetes, and (2) non-diabetes. The challenge of predicting heart disease was approached similar to supervised text classification assignment. Two degrees of discrimination were intended by the categorization task: (1) heart patients and (2) non-heart patients. RF, DT, BNB, GNB, G-BC, AB, M-LP, K-NN, Support Vector Machine, and Logistic Regression were among the ten machine-learning techniques utilized. we have proposed and developed our own CNN based model with 64 filters at convolution layer, 3 hidden layer with activation function Relu and sigmoid at final layer with pool size 2. The network was trained using 100 epochs with Adam optimizer using 10-fold cross-validation. Experiment 3, 4, 5, and 6 we have proposed and devloped LSTM, GRU, BILSTM, BIGRU with same hyperparmeters and parameter as CNN. Table 2 shows the parameters that was used for Deep Learning Methods.

Table 2 Hyperparameter Settings for Deep Learning Models

Hyperparameter	**Value**
Sequence Length	100
Batch Size	64
Learning Rate	0.001
Optimizer	Adam
Epochs	50
Loss Function	Binary Cross-Entropy
Activation Function	Softmax
Dropout Rate	0.5
Hidden Units (RNN Layers)	128
Kernel Size (CNN)	3x3
Stride (CNN)	1x1

In order to more accurately gauge the performance of machine learning algorithms, K-fold cross-validation was brought into action. For each experiment, K was set to a standard value of 10. The performance was reported using the weighted-average scores of Accuracy, Precision, Recall, and F_1.

5 Results and Analysis

Tables 1, 2, and 3 show the summarized results obtained by applying various machinelearning algorithms for Normal Diabetes Prediction, Diabetes during Pregnancy, and Heart Disease Prediction tasks respectively. The weighted-average F_1 scores were presented as a concluding measure for all tasks due to imbalanced datasets. The Table 1 shows, the best result with F_1 = 0.785292 using RF was obtained for Normal Diabetes Prediction. The performance comparison of all applied machine-learning algorithms for Normal Diabetes Prediction is evident from Table 1.

Table 3 Summarized Results of Normal Diabetes Prediction

ML Algorithm	F_1-Score
RF	**0.785291**
GNB	0.749944
G-BC	0.774546
LR	0.752286
M-LP	0.699831
DT	0.741640
K-NN	0.727619
SVC	0.626379
AB	0.745630
BNB	0.519567
CNN	0.513983
LSTM	0.526037
BILSTM	0.533086
GRU	0.578746
BIGRU	0.586426

Similarly, Table 2 shows, the best result with F_1 = 0.790618 using G-BC was obtained for Diabetes during the Pregnancy Prediction task. The performance comparison of all applied machine-learning algorithms for the Diabetes during Pregnancy Prediction task is clear from Table 2. Likewise, table 3 shows the best result with F_1 = 0.994 using LSTM was obtained for Heart Disease Prediction. The performance comparison of all applied 10

Table 4 Summarized Results on Diabetes during Pregnancy

ML Algorithm	F_1-Score
RF	0.780245
GNB	0.747802
G-BC	**0.790618**
LR	0.754727
M-LP	0.694850
DT	0.734672
K-NN	0.731603
SVC	0.608957
AB	0.750484
BNB	0.519567
CNN	0.521882
LSTM	0.528174

BILSTM	0.529484
GRU	0.580892
BIGRU	0.595866

machine-learning algorithms for Heart Disease Prediction can is demonstrated in Table 3.

Table 5 Summarized Results of Heart Disease Prediction

ML Algorithm	F_1-Score
RF	0.958783
GNB	0.826155
G-BC	0.937157
LR	0.837152
M-LP	0.821902
DT	0.945552
K-NN	0.711843
SVC	0.696894
AB	0.866180
BNB	0.826238
CNN	0.989846
LSTM	0.993246
BILSTM	**0.994382**
GRU	0.993246
BIGRU	0.993246

Among all machine learning algorithms, RF has shown outstanding performance for normal diabetes prediction, and heart disease prediction. The reason for the best performance of RF is due to following reasons.

5.1 RF

RF is an ensemble learning technique, combines multiple decision trees to make predictions. Each tree is trained on a different subset of the data, and the final prediction is determined by combining the individual tree predictions. This ensemble approach reduces over-fitting and improves the ability to generalize to new data. It also provides a measure of feature importance, indicating the relative significance of each feature in making accurate predictions. This information aids in identifying the most relevant features for prediction tasks, leading to better feature selection and engineering. By focusing on the most informative features, Random Forest enhances predictive performance. One of RF's strengths is its capability to capture nonlinear relationships between features and the target variable. Unlike linear models, Random Forest can model complex interactions and nonlinear patterns in the data. This flexibility is particularly advantageous for diabetes prediction, where the relationship between input features like glucose levels, BMI, and age, and the presence of diabetes may not follow a linear pattern. Random Forest exhibits robustness in handling outliers and missing data. The ensemble nature of the algorithm diminishes the influence of outliers on final predictions. Additionally, Random Forest can handle missing data by utilizing surrogate splits and imputing missing values based on other variables in the data set. This robustness allows for good performance even when the data has imperfections. To combat variance and over-fitting, Random Forest averages predictions from multiple trees. Each tree is trained on a different bootstrap sample of the data, and during the tree-building process, only a random subset of features is considered at each split. These randomization techniques reduce the correlation between individual trees and mitigate the risk of over-fitting. RF is also scale-able and efficient, making it suitable for large data sets with numerous features. The training process can be paralleled since the individual trees in the ensemble can be trained independently. This scalability and efficiency make Random Forest a practical choice for diabetes prediction, heart disease and other machine learning tasks, enabling faster processing and analysis. Table 2 shows, the best result with F_1 = 0.790618 using G-BC was obtained for Diabetes during the Pregnancy Prediction task among all machine learning algorithms. The possible reasons for the better

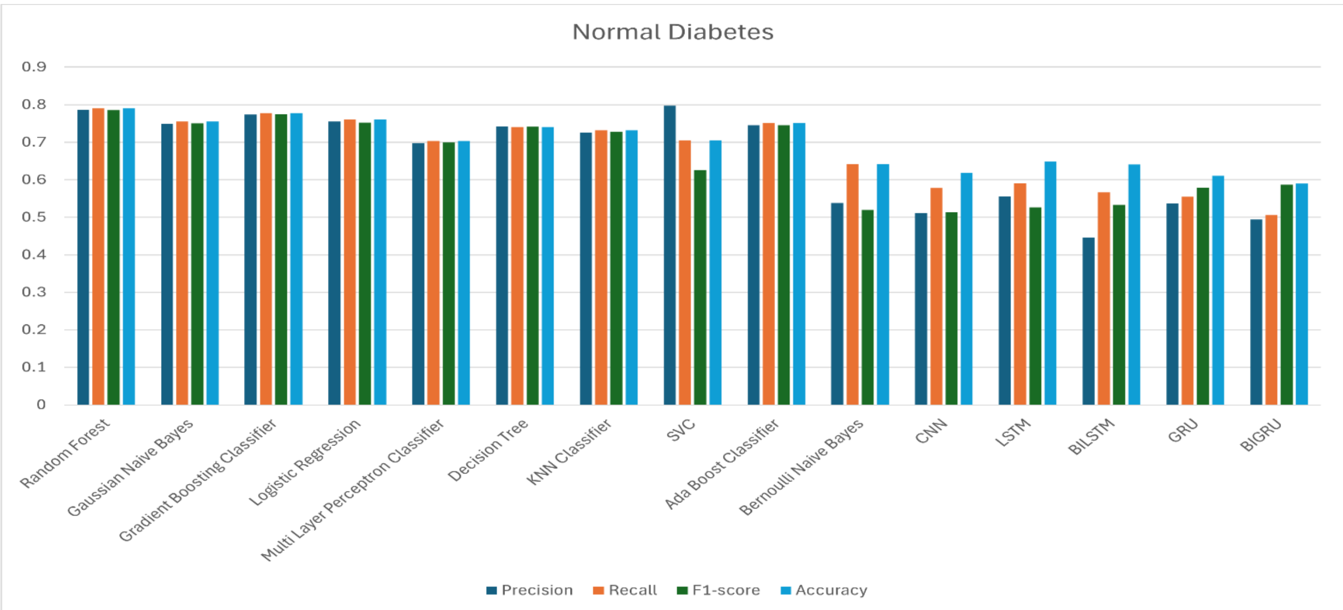

Figure 5 A performance comparison for Normal Diabetes Prediction tas

performance are as follows.

5.2 G-BC

G-BC combines multiple weak learners, such as decision trees, to create a strong predictive model. By adding learners sequentially to correct the mistakes of previous ones, GBC captures complex relationships in the data, enhancing predictive accuracy. G-BC utilizes a gradient descent optimization algorithm in training. By iteratively adjusting the model's parameters along the steepest descent of the gradient, GBC minimizes a loss function. This optimization method helps GBC find an optimal solution,reducing bias and variance and improving predictive accuracy.

It also effectively captures nonlinear relationships between features and the target variable, similar to Random Forest. It models complex interactions and nonlinearity in the data, which is advantageous for diabetes prediction, where the relationships between health indicators and diabetes presence can be nonlinear.

G-BC provides a measure of feature importance, allowing identification of the most relevant features for diabetes prediction. By focusing on these crucial features, GBC prioritizes and utilizes the informative aspects of the data, leading to improved predictions.

G-BC incorporates regularization techniques to prevent over-fitting and improve generalization performance. Methods like shrinkage/learning rate and feature subsampling control model complexity. These techniques reduce over-fitting and enable GBC to generalize well to unseen data, resulting in improved performance.

G-BC effectively handles imbalanced datasets, which are common in diabetes prediction and real-world applications. By assigning appropriate weights or using specialized loss functions, GBC gives more importance to the minority class (e.g., diabetes-positive cases). This ensures better predictive accuracy for both classes, addressing imbalanced data challenges.

Scalability and Efficiency: GBC is scalable and efficient, making it suitable for diabetes prediction with large data sets and numerous features. Optimized implementations paralleling the training process and utilize computational resources efficiently. This scalability allows GBC to handle complex diabetes prediction tasks effectively.

5.3 BILSTM

BiLSTM networks excel in predicting heart disease owing to their proficiency across several critical domains.

Primarily, they specialize in capturing prolonged dependencies within sequential data, pivotal for grasping the intricate interplay between historical health indicators and future risk over multiple time frames in the progression of heart disease. This capability is inherent in their architecture, which sustains a memory state over time. Moreover, BiLSTM networks process input sequences in both forward and backward directions, enabling them to assimilate context from past and future data points concurrently. This bidirectional approach is pivotal for capturing holistic patterns and dependencies, thereby enriching our comprehension of the evolution of heart disease over time.

Additionally, BiLSTM networks autonomously discern relevant features from input sequences during training. In the context of heart disease prediction, these features might encompass a wide array of physiological measurements such as heart rate, blood pressure, and cholesterol levels—key indicators of cardiovascular health.

Furthermore, BiLSTM networks demonstrate exceptional adaptability in handling input sequences of variable lengths, a crucial attribute in

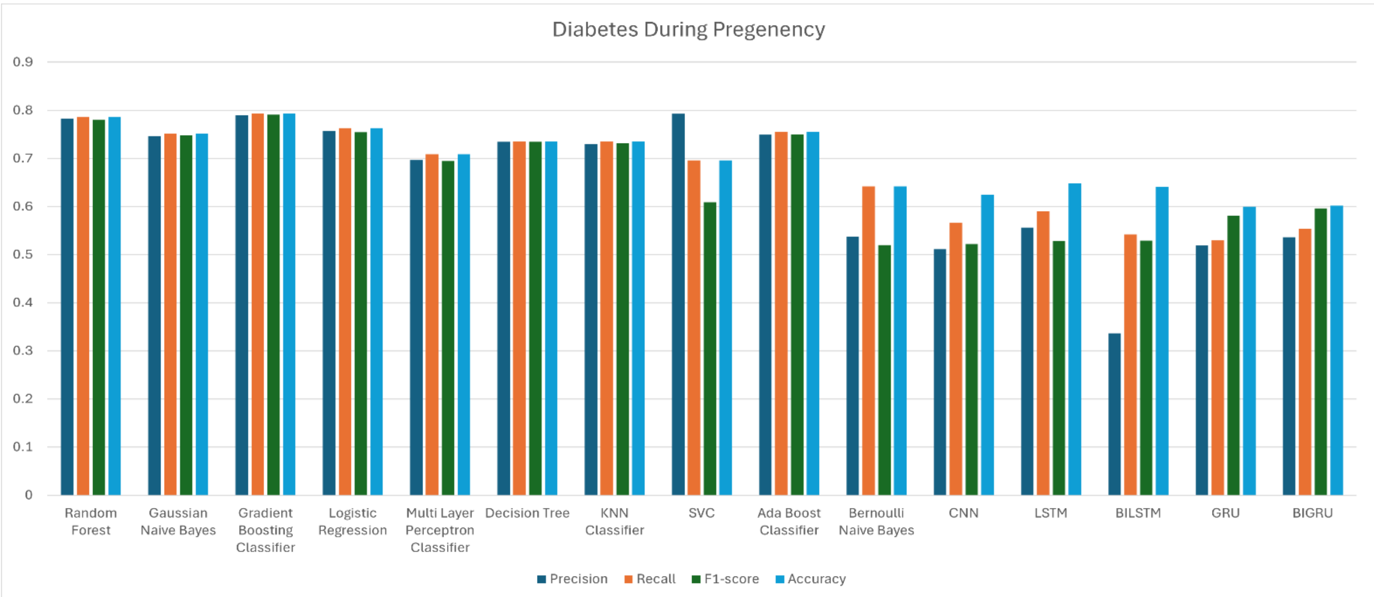

Figure 6 A performance comparison for Diabetes during Pregnancy Prediction task

healthcare data where the frequency and timing of health measurements often vary among patients. This adaptability empowers the model to accommodate diverse data formats and capture personalized disease progression patterns.
Lastly, BiLSTM networks exhibit robust learning capabilities even in scenarios with limited data—a common challenge in medical research due to the scarcity of largescale labeled datasets. Leveraging temporal dependencies within the data, BiLSTM networks efficiently utilize available information to make precise predictions

5.4 Best Methods

Our finding concludes that RF and G-BC and BILSTM are best suited for the classification tasks specially Diabetes Prediction and Heart Diabetes Prediction respectively. Figure 5 shows a detailed performance comparison for Normal Diabetes Prediction among all measures for all applied machine-learning algorithms.
Figure 6 shows a detailed performance comparison for Diabetes during the Pregnancy Prediction task among all measures for all applied machine-learning algorithms.
Figure 7 shows a detailed performance comparison for Heart Disease Prediction among all measures for all applied machine-learning algorithms.

5.5 Findings

Random Forest (RF) and BiLSTM performed best due to their respective strengths in handling feature importance and sequential data.
RF excels in handling diverse feature types and automatically identifies feature importance, making it highly effective for tasks like normal diabetes and heart disease prediction, where numerous variables contribute to the outcome. Its ensemble nature aggregates the results of multiple decision trees, which helps improve stability and robustness against overfitting.
BiLSTM, on the other hand, performs exceptionally well with sequential data, as it can capture long-term dependencies in the input sequences. This is particularly valuable for time-series or sequential prediction tasks like diabetes during pregnancy, where the relationship between past events and current outcomes is critical. By processing data in both forward and backward directions, BiLSTM enhances the model's ability to learn from past and future context, resulting in more accurate predictions.
Therefore, RF's success in handling feature importance and BiLSTM's ability to process sequential data contribute significantly to their top performances in the respective tasks.

6 Conclusion

A uniform approach to patient care, which can lead to an incorrect diagnosis, is often applied in healthcare, especially in diagnosis/prediction-related research, resulting in inadequate treatment and higher healthcare expenses. Machine learning methods in diabetes research can significantly improve diabetes research by effectively predicting diabetes, identifying risk factors, and providing personalized treatment. The study proposes three novel benchmark corpora based on real cases of heart patient records, normal diabetes patients, and diabetes during pregnancy, respectively. As a major contribution, all of these corpora were evaluated using 10 different machine learning algorithms, including RF, GNB, G-BC, Logistic Regression, M-LP, DT, KNN Classifier, SVC, Ada Boost Classifier, and BNB. The study provides a detailed and in-depth comparison of

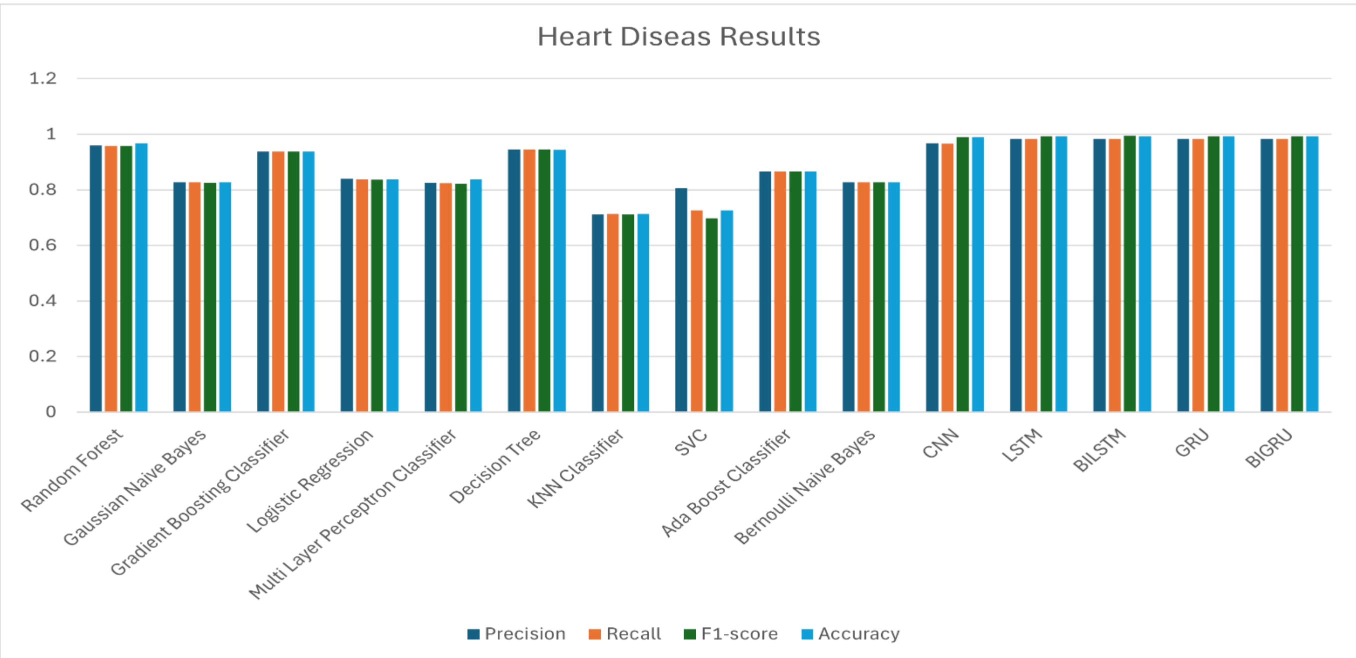

Figure 7 A performance comparison for Heart Disease Prediction task

applied machine learning algorithms for high-risk disease prediction, such as Heart Disease Prediction, Normal Diabetes Prediction, and Diabetes during Pregnancy Prediction tasks. The findings from this research have the potential to revolutionize healthcare practices by enabling more accurate, personalized, and cost-effective disease predictions, ultimately improving patient outcomes and reducing healthcare costs.

7 Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

[1] http://www.ishp.gov.al/wpcontent/uploads/205/kalendar

[2] https://www.familjadheshendeti.com/semundja-e-sheqerit-diabeti-te-femrat/S.

[3] Bo He, Kuang-i Shu, and Heng Zhang, "Machine Learning and Data Mining in Diabetes Diagnosis and Treatment," IOP Conference Series: Materials Science and Engineering, Volume 490, Issue 4, IOP Conf. Series: Materials Science and Engineering 490 (2019) 042049 IOP, doi:10.1088/1757899X/490/4/042049.

[4] Llaha, Olta, and Amarildo Rista. "Prediction and Detection of Diabetes using Machine Learning."

[5] Singh, Archana, and Rakesh Kumar. "Heart disease prediction using machine learning algorithms." 2020 International Conference on Electrical and Electronics Engineering (ICE3). IEEE, 2020.

[6] Pal, Saurabh. "Chronic Kidney Disease Prediction Using Machine Learning Techniques." Biomedical Materials and Devices (2022): 1-7.

[7] Lukmanto, Rian Budi, Ariadi Nugroho, and Habibullah Akbar. "Early detection of diabetes mellitus using feature selection and fuzzy support vector machine." Procedia Computer Science 157 (2019): 46-54.

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

[9] Chelaramani, Sahil, et al. "Multi-task knowledge distillation for eye disease prediction." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021.

[10] Y. HaCohen-Kerner, Z. Gross, and A. Masa, "Automatic extraction and learning of keyphrases from scientific articles," In International Conference on Intelligent Text Processing and Computational Linguistic, vol. 3406, no. February 2005, pp. 657–669, 2005, doi: 10.1007/978-3-540-30586-6-74.

[11] N. Pendar and E. Cotos, "Automatic identification of discourse moves in scientific article introductions," no. June, pp. 62–70, 2008, doi: 10.3115/1631836.1631844.

[12] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," Lr. 2018 11th Int. Conf. Lang. Resour. Eval., pp. 3483–3487, 2019.

[13] RTA-CSIT Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... Virani, S. S. (2019). Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation, 139(10), e56-e528. doi: 10.1161/CIR.0000000000000659.

[14] American Diabetes Association. (2021). Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2021. Diabetes Care, 44(Supplement 1), S15-S33. doi: 10.2337/dc21-S002.

[15] Fletcher, B. R., Hinton, L., Hartmann-Boyce, J., Roberts, N. W. (2016). A systematic review of healthcare providers' and patients' diabetes attitudes, perceptions, and expectations and how these influence patient outcomes. Diabetes Research and Clinical Practice, 113, 1-8. doi: 10.1016/j.diabres.2016.01.024.

[16] Lomotan, E. A., Hoeksel, R. (2018). Leveraging health information technology to improve management of chronic conditions in underserved populations. Current Diabetes Reports, 18(6), 32. doi: 10.1007/s11892-018-1009-3.

[17] Mishra, S. R., Neupane, D., Bhandari, P. M. (2019). Health information systems in developing countries: Challenges in implementing and sustaining the impact. Global Health Action, 12(1), 1559497. doi: 10.1080/16549716.2018.1559497.

[18] Patel, A., Raju, D. (2019). Use of health information technology in chronic disease management: A literature review. International Journal of Medical Informatics, 130, 103941. doi: 10.1016/j.ijmedinf.2019.07.017.