SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V9I12P115

Volume 9 Issue 12, 167-179, December 2022

Original Article

Diabetes Prediction using Improved Artificial Neural Network using Multilayer Perceptron

T. Madhubala[1], R. Umagandhi[2], P. Sathiamurthi[3]

[1]*Computer Science, Kongunadu Arts and Science College, Coimbatore, Tamil Nadu, India.*
[2]*Department of Software Systems, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India.*
[3]*Department of Mechanical Engineering, Kongu Engineering College, Perundurai, Tamil Nadu, India.*

[1]*Corresponding Author : madhumadhi34@gmail.com*

Received: 25 October 2022 Revised: 29 November 2022 Accepted: 14 December 2022 Published: 25 December 2022

Abstract *- Diabetes is a continual health situation that involves how your corpse revolves foodstuff into power. Our body severs most of the food we eat into glucose and discharges it into our bloodstream. When our blood sugar goes up, it signs our pancreas to liberate the insulin. Insulin takes steps to permit the blood glucose into our body's cells for use as liveliness. If the patient has diabetes, the body cannot produce sufficient insulin. Unmanaged diabetes comes due to hunger, less-healing wounds, frequent urination, more thirst, fatigue, itchy skin, dry and blurry revelation, etc. Hence detection of diabetes is a very virtual role in our country and abroad. The proposed system is Improved Artificial Neutral Network (IANN), which used Pima Indian Diabetes (PID) dataset to construct the deep network. Also, this research used the Weka tool to find the existing classifier performance and proposed IANN outperforms compared to other algorithms.*

Keywords *- Data Mining, Machine Learning, Artificial Neural Network, Diabetics and Multilayer Perceptron.*

1. Introduction

Diabetes is a foremost community health problem, evolving as a pandemic. An essential reason for fatness and bulky weight is a body health inequity between calories spent and charred up. Substantial activity is the unique backbone for preventing diabetes mellitus. Decreasing body weight, monitoring good health, decreasing energy and food intake, and taking more fruits, vegetables, legumes, whole grains, nuts and crops are vital for human life [1]. Rising indication proposes that healthcare guidance can decrease alcohol routinely, raise physical action, and progress the acceptance of further needed lifestyle behaviors among patients. Even though passing involvements and patient dealings can nature, the increasing embracing of multiple changeable behaviors is less healthy and considered for diabetes disease control and prevention. This survey focused on patient interactions that can positively impact prediabetes and motivate the patients to modify their lifestyle behavior to reduce diabetes [2].

1.1 Type of Diabetes

Identified as Type 1 diabetes is attention to be an autoimmune situation which attacks the mistakenly in your immune system. Due to this, beta cells are damaged permanently in the pancreas, which cannot produce insulin. Incorrectly violent and obliterates the beta cells in the pancreas that make insulin. The destroying the cells are permanent. This kind of diabetes comes from genetic, environmental and lifestyle factors.

Type 2 diabetes begins when the occurrence of insulin resistance. Our body dismisses the insulin, and the pancreas yields high insulin to unable to stock the further use and which produces more blood sugar. Because of this situation, insulin production is very lower. This type of diabetes comes from genetics, more weight and obesity and a highly sedentary lifestyle.

Gestational diabetes is caused by insulin-blocking hormones, which happen during pregnancy. It is named prediabetes, and a family history of diabetes. About 50 % of trusted sources of people make a diagnosis of gestational diabetes.

1.2. Diabetes Care

Care of diabetes is very important. When sugar can is near standard, they get more energy to the body, are not tired and thirsty, pass urine less often, have fewer bladder infections and less casual of stroke, heart attack, and eye problems. Get relief from the problem of kidney, teeth and gum, eye and nerve damage and heal better. For removing diabetes, the diabetes meal plan help, which is framed into the health care team. Eat foods lower in calories, salt and sugar. Do not use regular soda and juice. Drink water and eat foods that look like vegetables, fruits, whole grains, bread and cereals. When eating a meal, wheat pasta, brown rice, fruits and vegetables, and beans can add to your food. Figure 1 shows the body components related to diabetics, and Figure 2 represent the usage of food for diabetic patient.

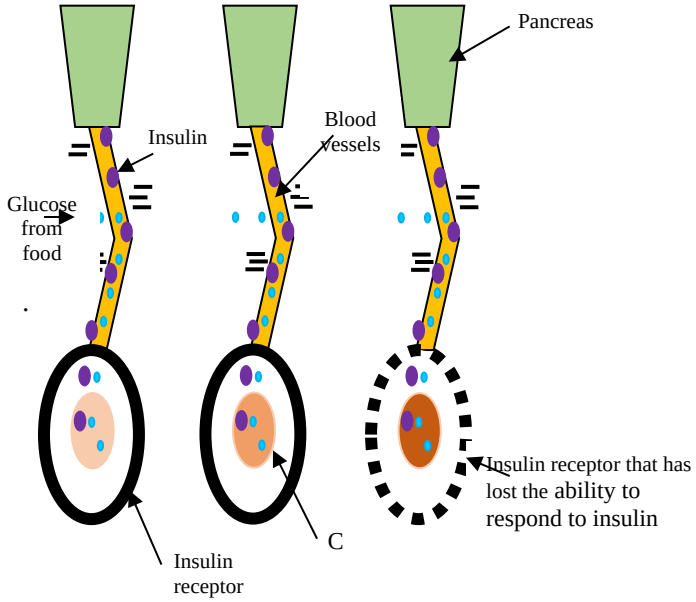

Fig. 1 Functional components in the human body

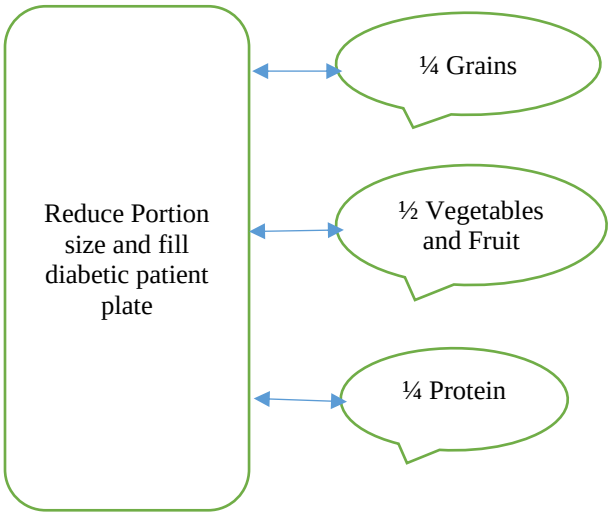

Fig. 2 Usage of food for diabetic patients.

2. Related Work

In the United States have diabetes for 37.3 million trusted people. Five to ten percentages of people have type 1 diabetes, and ninety to ninety-five percentages of people are affected by type 2 diabetes. Handling the major disease of diabetes in elder persons can be inspiring due to combining many diseases and age-associated incapacities, specifically from the perspective of insulin treatment. Authors [1] reviewed insulin usage on the healthcare side. Even though the pervasiveness and difficulty of giving elder persons with diabetes, there is an amazing absence of suggestions about how persons handle their situations. The authors [2-5] focused on insulin treatment-related problems. Jenifer Florence et al. [36] focused on the diet, treatment and excise level of diabetic patients and their depression of the patients. They proposed mixed-method learning in a rural greater than 30 years. The authors used the Wilcoxon rank test to analyse the data. These sensitive techniques can decrease complications and produce social support for diabetic patients.

Xiangli et al. [43] reviewed the Glucagon-like peptide -1 receptor agonists, irrespective of their work of different generic parts, major aids in decreasing the danger of the complexity, particularly in innovative macro albuminuria related to type two diabetes patients. AurelieBerot et al. [8] focused on type 1 diabetes disease with the help of French health insurance data over 2015. These works describe their medical performance from 1997 to 2019 at Reims University. The Champagne-Ardenne Diabetes Network database mines scientific and natural data. AndapetronelaRadan et al. [9] mobilized pregnant women with pre-existing diabetes. Also extant with added severe illness, as associated with non-diabetic focuses. Now, the recommendation is limited regarding the responsibility of GD in the severity of Covaid-2 disease during pregnancy. The authors applied the logistic regression of machine learning algorithms and demonstrated that 34.6% of the patients had gestational diabetes.

Gallagher et al. [41-42] surveyed about HbA1c.HbA1c has been the same as plasma glucose. Due to this, retinopathy grows rapidly [12]. A recent report from Australia, the world health organization [12-14], says the HbA1c test is better for predicting retinopathy and microvascular complications. The HbA1c can analysis by the diabetes team suggested that HbA1c is employed in the study of diabetes, and research can make when the HbA1c point is ≥6.5% and otherwise taken as intermediate hyperglycemia is present [38]. HbA1c imitate the blood glucose meditation in the earlier three months and gives normal and stable hemoglobin. Hence this test is very powerful in imperfect resource places [16-21].

Rimesh Pal et al. [39] experimented with the diabetic ketoacidosis illness hyperglycemic hyperosmolar syndrome for COVID-19 patients by diagnosing it with biochemical parameters. The authors determined that the pH of patients who died and were discharged from hospitals was lower. The Covid-19 patients with people with diabetes are analyzed in [23-26].

3. Materials and Methods

Before loading into the dataset in the Weka tool, the dataset contains the following structure exposed in Figure 3.

First, upload the Pima Indian dataset into the python Weka tool with the help of open file options. Figure 4 represents the Minimum, Maximum, Mean statistics and Standard Deviation for the Pima Indian dataset.

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
3,78,50,32,88,31.0,0.248,26,1
10,115,0,0,0,35.3,0.134,29,0
2,197,70,45,543, 30.5,0.158,53,1
8,125,96,0,0,0.0,0.232,54,1
4,110,92,0,0,37.6,0.191, 30,0
10,168,74,0,0,38.0,0.537,34,1
10,139,80,0,0,27.1,1.441,57,0
1,189,60,23,846, 30.1,0.398,59,1
5,166,72,19,175,25.8,0.587,51,1
7,100,0,0,0,30.0,0.484,32,1
0,118,84,47,230,45.8,0.551,31,1
7,107,74,0,0,29.6,0.254,31,1
1,103,30,38,83,43.3,0.183,33,0
1,115,70,30,96,34.6,0.529,32,1
3,126,88,41,235, 39.3,0.704,27,0
8,99,84,0,0,35.4,0.388,50,0

Fig. 3 Pima dataset unstructured manner

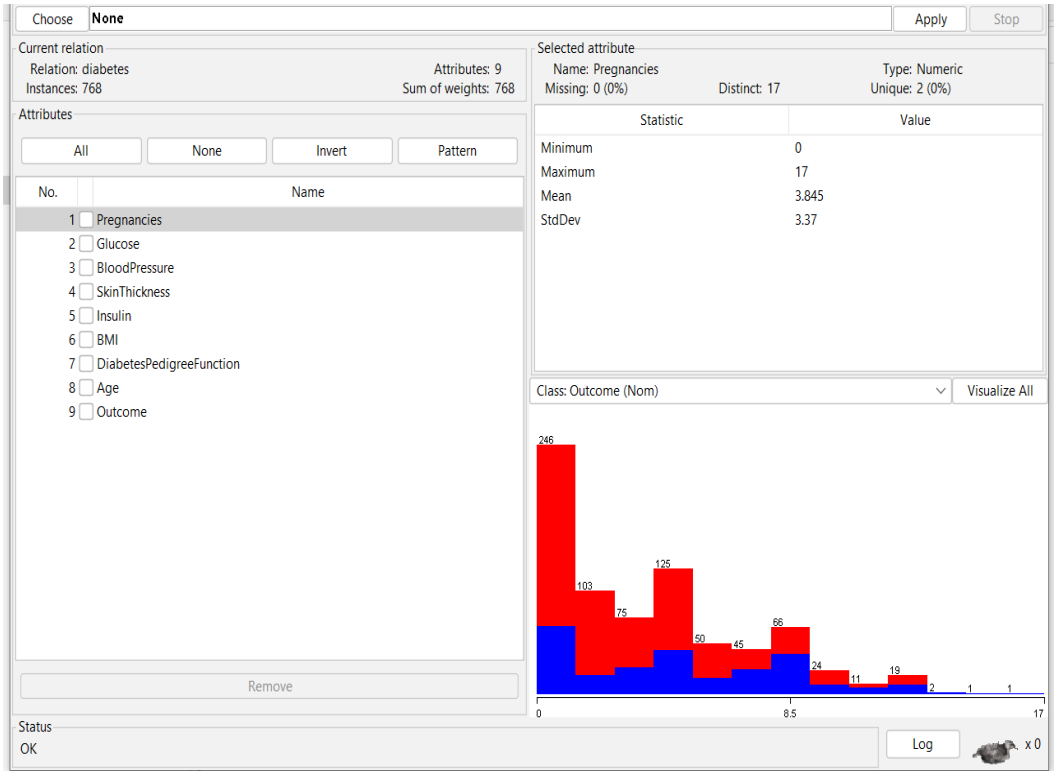

Fig. 4 Loading dataset

After loading the dataset, you can click the edit menu in Wekatool; it represents the content of the data in Figure 5.

Relation: diabetes

No.	1: Pregnancies Numeric	2: Glucose Numeric	3: BloodPressure Numeric	4: SkinThickness Numeric	5: Insulin Numeric	6: BMI Numeric	7: DiabetesPedigreeFunction Numeric	8: Age Numeric	9: **Outcome** Nominal
1	6.0	148.0	72.0	35.0	0.0	33.6	0.627	50.0	yes
2	1.0	85.0	66.0	29.0	0.0	26.6	0.351	31.0	no
3	8.0	183.0	64.0	0.0	0.0	23.3	0.672	32.0	yes
4	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21.0	no
5	0.0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	yes
6	5.0	116.0	74.0	0.0	0.0	25.6	0.201	30.0	no
7	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26.0	yes
8	10.0	115.0	0.0	0.0	0.0	35.3	0.134	29.0	no
9	2.0	197.0	7	Right click (or left+alt) for context menu			0.158	53.0	yes
10	8.0	125.0	96.0	0.0	0.0	0.0	0.232	54.0	yes
11	4.0	110.0	92.0	0.0	0.0	37.6	0.191	30.0	no
12	10.0	168.0	74.0	0.0	0.0	38.0	0.537	34.0	yes
13	10.0	139.0	80.0	0.0	0.0	27.1	1.441	57.0	no
14	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59.0	yes
15	5.0	166.0	72.0	19.0	175.0	25.8	0.587	51.0	yes
16	7.0	100.0	0.0	0.0	0.0	30.0	0.484	32.0	yes
17	0.0	118.0	84.0	47.0	230.0	45.8	0.551	31.0	yes
18	7.0	107.0	74.0	0.0	0.0	29.6	0.254	31.0	yes
19	1.0	103.0	30.0	38.0	83.0	43.3	0.183	33.0	no
20	1.0	115.0	70.0	30.0	96.0	34.6	0.529	32.0	yes
21	3.0	126.0	88.0	41.0	235.0	39.3	0.704	27.0	no
22	8.0	99.0	84.0	0.0	0.0	35.4	0.388	50.0	no
23	7.0	196.0	90.0	0.0	0.0	39.8	0.451	41.0	yes
24	9.0	119.0	80.0	35.0	0.0	29.0	0.263	29.0	yes

Add instance

Fig. 5 before Normalization

Viewer

Relation: diabetes-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0

No.	1: Pregnancies Numeric	2: Glucose Numeric	3: BloodPressure Numeric	4: SkinThickness Numeric	5: Insulin Numeric	6: BMI Numeric	7: DiabetesPedigreeFunction Numeric	8: Age Numeric	9: **Outcome** Nominal
688	0.05882352941...	0.5376884...	0.409836065573...	0.191919191919...	0.0	0.4217...	0.0439795046968403	0.1333...	no
689	0.05882352941...	0.7035175...	0.606557377049...	0.262626262626...	0.212765...	0.3591...	0.3202391118701964	0.0333...	no
690	0.05882352941...	0.7236180...	0.672131147540...	0.464646464646...	0.212765...	0.6870...	0.10973526900085397	0.4166...	yes
691	0.47058823529...	0.5376884...	0.655737704918...	0.0	0.0	0.3666...	0.33219470538001705	0.2166...	no
692	0.76470588235...	0.7939698...	0.934426229508...	0.0	0.0	0.6304...	0.07643040136635354	0.3833...	yes
693	0.11764705882...	0.6080402...	0.573770491803...	0.323232323232...	0.112293...	0.5827...	0.34500426985482496	0.0333...	no
694	0.41176470588...	0.6482412...	0.557377049180...	0.494949494949...	0.147754...	0.5737...	0.15414175918018785	0.3666...	yes
695	0.11764705882...	0.4522613...	0.491803278688...	0.0	0.0	0.3502...	0.04824935952177626	0.0666...	no
696	0.41176470588...	0.7135678...	0.737704918032...	0.242424242424...	0.567375...	0.4530...	0.02134927412467976	0.3666...	yes
697	0.17647058823...	0.8492462...	0.606557377049...	0.191919191919...	0.147754...	0.4456...	0.08112724167378309	0.1666...	yes
698	0.0	0.4974874...	0.0	0.0	0.0	0.3725...	0.07472245943637916	0.0166...	no
699	0.23529411764...	0.6381909...	0.721311475409...	0.111111111111...	0.183215...	0.5141...	0.2220324508966695	0.1166...	no
700	0.23529411764...	0.5929648...	0.573770491803...	0.0	0.0	0.6631...	0.35269000853970967	0.0833...	no
701	0.11764705882...	0.6130653...	0.622950819672...	0.272727272727...	0.236406...	0.5350...	0.17292912040990605	0.0833...	no
702	0.35294117647...	0.6281407...	0.639344262295...	0.313131313131...	0.0	0.4113...	0.20794192997438082	0.4666...	yes
703	0.05882352941...	0.8442211...	0.721311475409...	0.292929292929...	0.0	0.5216...	0.35311699402220326	0.5166...	yes
704	0.11764705882...	0.6482412...	0.0	0.0	0.0	0.5737...	0.0964987190435525	0.3333...	no
705	0.23529411764...	0.5527638...	0.622950819672...	0.202020202020...	0.118203...	0.4232...	0.017079419299743805	0.1	no
706	0.35294117647...	0.4020100...	0.655737704918...	0.363636363636...	0.0	0.5931...	0.042271562766865924	0.1166...	no
707	0.58823529411...	0.5778894...	0.0	0.0	0.0	0.0	0.07813834329632792	0.15	yes
708	0.11764705882...	0.6381909...	0.377049180327...	0.212121212121...	0.395981...	0.5126...	0.041844577284372325	0.0166...	no
709	0.52941176470...	0.8241206...	0.639344262295...	0.0	0.0	0.4888...	0.02988898377455166	0.4	yes
710	0.11764705882...	0.4673366...	0.524590163934...	0.323232323232...	0.189125...	0.5663...	0.2544833475661828	0.0333...	yes
711	0.17647058823...	0.7939698...	0.524590163934...	0.131313131313...	0.457446...	0.4649...	0.09265584970111014	0.05	no

Add instance

Fig. 6 Content of PID dataset after normalization applied

After loading the dataset, the pre-processing stage is important. For any missing value, replace the lost value from the Weka filtering. In this stage, inconsistent and incorrect data are removed with the help of replacing missing values and normalization from the Weka filters. No missing value in the dataset. Hence normalization is applied for values lying from zero to one. This is shown in below Figure 6.

The next step is to build the classifier model with the help of classify tab. For training the classifier, the following classifiers are used.

3.1. IBk Algorithm

IBk (Instance-Based learner based on k neighbor) is the nearest neighbor approach; it uses distance measure and is named a lazy learner algorithm. They explicitly gave the number of nearest neighbors. Various algorithms can be used to determine the nearest neighbor quickly. When the time of classification, this algorithm provisions the dataset and performs an action on the PID data instead of learning the training set immediately.

 # Instance Based Learner algorithm

Set of training sample DB
Let w1…. Wk represents the k instances from D
Nearest to w_q return
Give the new data points to the group (Neighbor maximum)
Given query instance w_q to be classified.

3.2. J-48 Decision Tree

The J48 decision tree on WEKA uses the Iterative Dichotomiser 3 algorithm. Internal nodes of this decision tree represent the various features, and a branch indicates the observed samples. Finally, terminal nodes are the dependent variable for predicted attributes, and other features are the independent variables in the dataset. J48 is the operation of the open-source C4.5 algorithm in Weka. It generates the rules from Pima diabetic's data. The training data is a prior classified sample of set Z= {z1, z2,}. Each sample zi contains an m-dimensional vector (y1, i, y2, i… ym, i) where the yj shows the attributes for the corresponding sample. The greatest information gain produces more classification accuracy.

Steps to implement the J-48

- When the instances belong to a similar class, the leaf is labelled as a similar class
- Information gain is measured for each feature from the test on the features
- Based on the currently chosen parameter, the best feature is taken

3.3 OneR Algorithm

This algorithm learns a one-level decision tree and creates a group of rules which test the specific nominal feature.

Steps

- One division for each of the values of the features
- Each division gives the most repeated class
- Error rate: fraction of instances that don't fit into the popular class of their equivalent division
- Select the attributes with the lowest error rate

3.4. Multi-Layer Perceptron (MLP)

Multilayer perceptron uses backpropagation to learn the classify instances. The network parameters are modified throughout the training time and contain unthresholded linear units of all sigmoids. Between nodes of initial weight is adjusted, and the splitting data is shuffled with seed help. The option Momentum means applied to the weight updates—the classifier throughput when nominal features in the dataset are increased to nominalToBinary pre-processing filters. In the neural network, hidden Layers are positive whole numbers, one indicates the presence of hidden layers, and zero means no hidden layers. A simple structure of Multi-Layer Perceptron and layer generations in Weka tools are exposed in Figure 7 and Figure 8, respectively.

3.5. Random Forest

A random forest is a Machine Learning (ML) approach utilized to resolve regression and classification problems. It uses ensemble learning and associates more classifiers to clarify multifaceted problems.
Implementation steps:

- Choose random K data points from the training set.
- Construct the decision trees associated with the chosen Subsets.
- ChooseN for decision trees to build and repeat the random K selection and build the decision trees.
- For new data points, determine the predictions of each DT, and consign the new data points to the group that succeeds in the popular votes.

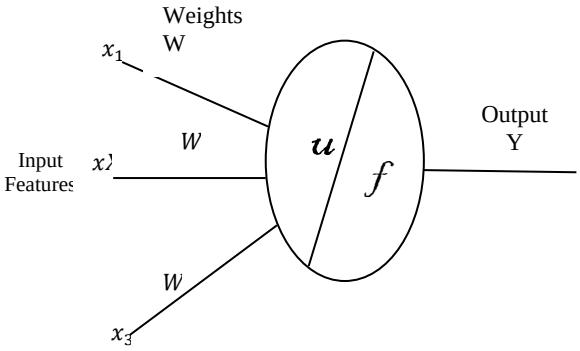

Fig. 7 Structure of the Mult-Layer Perceptron

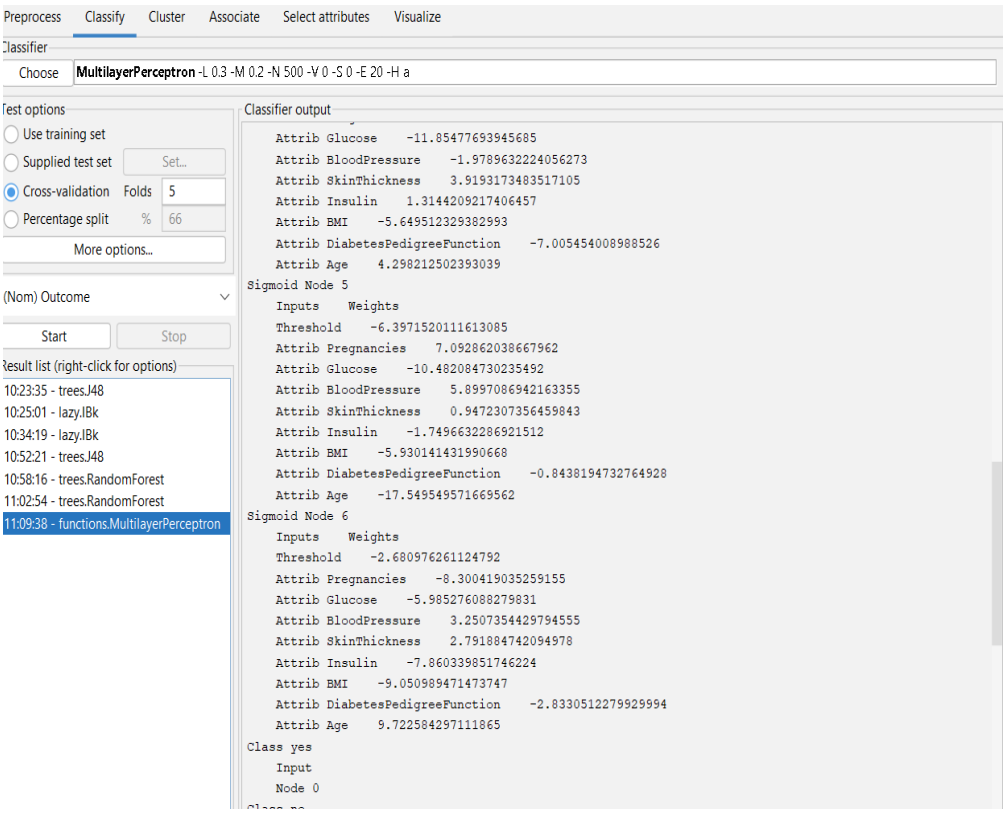

Fig. 8 Layers generation in MLP

3.5. NavieBayes Algorithm

A Naive Bayes classifier is called a probabilistic ML model employed for classification jobs. The Bayes in Naïve Bayes comes from Bayes' Theorem. Bayes' Theorem defines the probability of an occurrence according to prior knowledge related to the event. Naive Bayes is a linear classifier. Naive Bayes used alpha and beta hyperparameters to adjust for smoothing out. Due to this process, the training set can optimize.

Implementation steps:

- Separate using class labels.
- Review the dataset and data by class
- Apply probability distribution function
- Class Probabilities.

3.6. Proposed Improved Artificial Neural Network

This proposed research uses the PID dataset to construct the neural network. Python contains numerous build libraries. Panda's library is used to load the PID dataset. For mathematical calculation, Numpy is used. The keras model is used to construct the deep neural network. After importing libraries and loading the dataset, we need to separate the input and target fields. The proposed model flow and architecture are shown in Figure 9 and Figure 10.

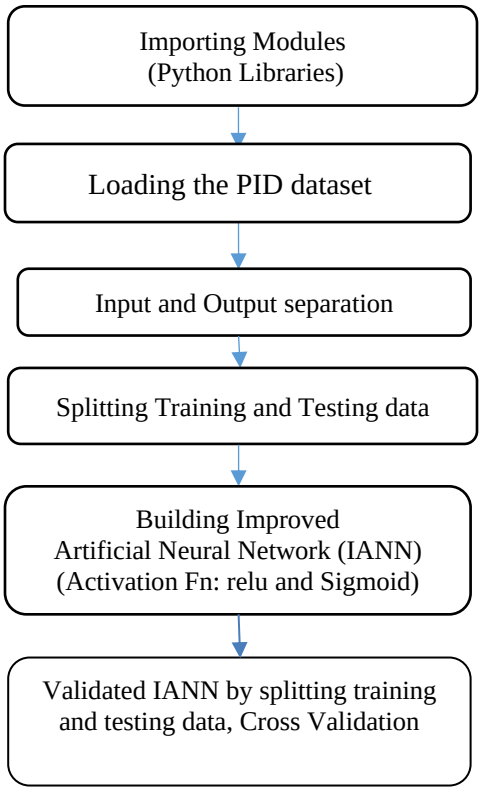

Fig. 9 Proposed Improved Artificial Neural Network

Data may be separated into training and testing data to build the classification model. In model building, activation functions are essential features in artificial neural networks (ANN). This research used ReLU and Sigmoid functions to construct the model. In ANN, the activation function describes the output of that node specified as an input set. Tensor flow is the framework to implement the deep learning model.

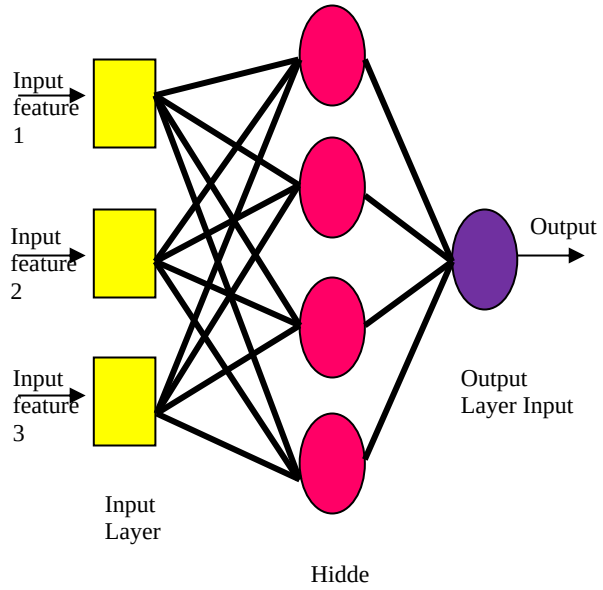

Fig. 10 Layer description

Table 1. Algorithm pseudo code to find Training accuracy and Testing Accuracy

```
#import the python libraries
        Numpy, Pandas, tensorflow, Keras and Sklearn
Loading the diabetes dataset
# Separating the input and output attributes
        Predictor = df.iloc[:, :-1]
        output = df.iloc[:,8]
# Split the training and testing data
        W_train_A, Z_test, Z_train_B, Z_test =
        train_test_split(Predictor, output,
        random_state=42)
        W_train, W_valid, Z_train, Z_valid =
        train_test_split(W_train_A, Z_train_B,
        random_state=42)
# Model creation
Numpy. random.seed(7)
tensorFlow.random.set_seed(7)
modeliann =Sequential()
modeliann.add(Dense(15,input_dim=8,
activation='relu'))
modelann.add(Dense(10,activation=relu'))
modeliann.add(Dense(8,activation='relu'))
modeliann.add(Dropout(0.25))
modeliann.add(Dense(1, activation='sigmoid'))
Find Training accuracy and Testing Accuracy
```

Table 2. Algorithm pseudo code to find Training accuracy and Testing Accuracy

```
#import the python libraries
        Numpy, Pandas, tensorflow, Keras and Sklearn
Loading the diabetes dataset
# Converting the dataframe into an array
    Predictor = df.iloc[:, :-1]
    Target = df.iloc[:,8]
cvscoress = []
    W=np.array(Predictor)
    Z=np.array(Target)
# Cross validation
fkfold = StratifiedKFold(n_splits=10, shuffle=True,
random_state=7)
# Model creation
for ftrain,ftest in fkfold.split(W, Z):
modeliann=Sequential()
modeliann.add(Dense(15,input_dim=8,
activation='relu'))
modeliann.add(Dense(10,activation='relu'))
modeliann.add(Dense(8,activation='relu'))
modeliann.add(Dropout(0.25))
modeliann.add(Dense(1, activation='sigmoid'))
#Compile the model
Modeliann.compile(loss="binary_crossentropy",
Optimizer="SGD",
Metrics=['accuracy'])
modeliann.fit(W[ftrain], Z[ftrain], epochs=20)
   scores = modeiann.evaluate(W[ftest], Z[ftest],
verbose=0)
   print("%s: %.2f%%" % (modeliann.metrics_names[1],
scores[1]*100))
cvscoress.append(scores[1] * 100)
```

The proposed architecture is a modified MLP, which uses four hidden layers. Sigmoid and Relu activation functions are used.

Table 1 shows the algorithm pseudo code to find the proposed work's training and testing accuracy. Step 1 Imported the python libraries and loaded the PID dataset. After loading, separate the train data and test data. Then proposed model is created with four hidden layers and epochs is 200. Finally, the user gets the accuracy of the classifier. Table 2 shows the algorithm pseudo-code for cross-validation measures. Here K-fold, split=10 and epochs=20 are used.

4. Experimentation and Results

Experiments can be carried out using the python Weka machine-learning tool. Then test the performance of the different classifiers derived from machine learning.

4.1. Dataset Description

The proposed techniques emphases on diabetic patients in instructing them to deliver a sufficient analysis. Pima

Indians Diabetes (PID) is a benchmark dataset from the UCI ML repository [30-32]. It is used as training and test data for predicting the model. This data contains patients who are females at least 21 years old of Pima Indian heritage. Every suffering from diabetics, disease person is sharp through 8 descriptors and 2 classes. The patient proves signs of diabetes based on the World Health Organization (WHO) norms. Table 3 and Table 4 represent the information about the PID benchmark data.

Table 3. Features of the Pima Indian diabetes

No. of Features	Name of the feature	Descriptions
W1	Pregnancies for Women	Number of times pregnant
W2	Glucose Level	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
W3	Blood Pressure	Diastolic blood pressure (mm Hg)
W4	Skin Thickness	Triceps skin fold thickness (mm)
W5	Insulin level	2-Hour serum insulin (mu U/ml)
W6	BMI	Body mass index (weight in kg/(height in m)^2)
W7	Diabetes Pedigree Function	Diabetes pedigree function
W8	Age of patient	In years
Z	Outcome-class variable	0 or 1

Table 4. Parameter of the Pima Indians Diabetes benchmark dataset

Dataset	Samples	Descriptors	Number of features	Classes
Pima Indians diabetes	768	8	9	2

4.2 Measure Used

4.2.1. Precision

Proportion of appropriately identified positive interpretations to all expected positive observations as shown below Equation (1).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (1)$$

4.2.2. Recall

Proportion of appropriately forecasted positive interpretations to the total observation in the actual class as represented in Equation (2).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (2)$$

Here TP: True Positive, FP: False Positive, TN: True Negative and FN: False Negative

4.2.3. F-Measure

Weighted average of precision and Recall false positives and false negatives are considered. It is given by Equation (3).

$$\text{F} - \text{Measure:} \frac{2(\text{Recall} * \text{Precession})}{(\text{Recall} + \text{Precission})} \qquad (3)$$

4.3. Experiment Result and Discussion

The experiment result shows different data mining results of classifier measures for NaiveBayes (NB), MultiLayer Perceptron (MLP), OneR, J48 IBk, J48, OneR, MultiLayer Perceptron (MLP), RandomForest (RF), MultiLayer Perceptron (MLP), and NaiveBayes (NB) algorithms using Weka for 5-fold and 10-fold measures of cross-validation, which are shown in Figure 11 and Figure 12 respectively. NaieBayes makes superior classification outcomes and is compared to other algorithms in the Pima diabetic dataset. The throughput of the classifier using 5-fold cross-validation is represented in Table 5. NaiveBayes produces 76.30% of the data correctly classified among all the classifiers. Similarly, from available classifiers, NB produces a better result of 76.30% as a correctly classified instance among all the classifiers in the class label as an outcome in the PID dataset using the cross-validation of the 10-fold test, which is represented in Table 6

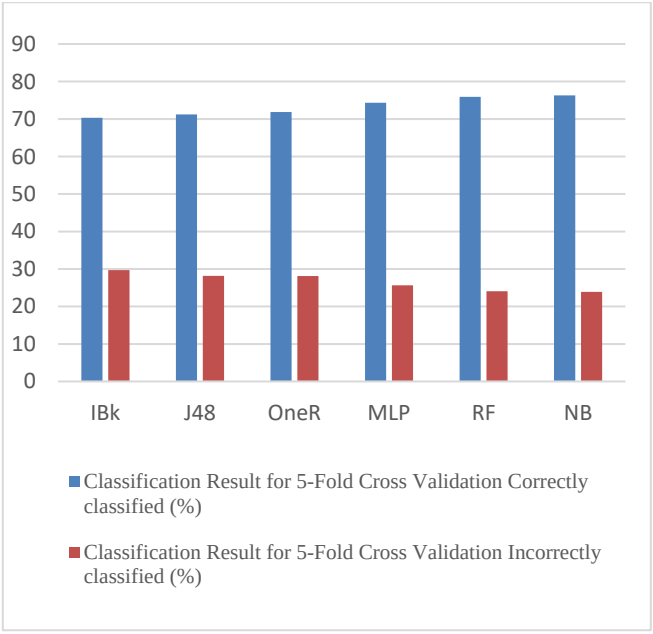

Fig. 11 Classifiers - 5-Fold Cross Validation

Table 5. Correctly and incorrectly classification using 5-Fold Cross Validation

Classifiers	Correctly Classified Instance (%)	Incorrectly Classified Instance (%)
IBk	70.31	29.68
J48	71.22	28.17
OneR	71.87	28.12
MultiLayer Perceptron (MLP)	74.34	25.65
RandomForest (RF)	75.91	24.08
NaiveBayes (NB)	76.30	23.9

Table 6. Correctly and incorrectly classification using 10-Fold Cross Validation

Classifiers	Correctly Classified Instance (%)	Incorrectly Classified Instance (%)
IBk	70.18	29.81
OneR	70.83	29.16
J48	73.82	26.17
MLP	75.13	24.86
RF	75.91	24.08
NB	76.30	23.69

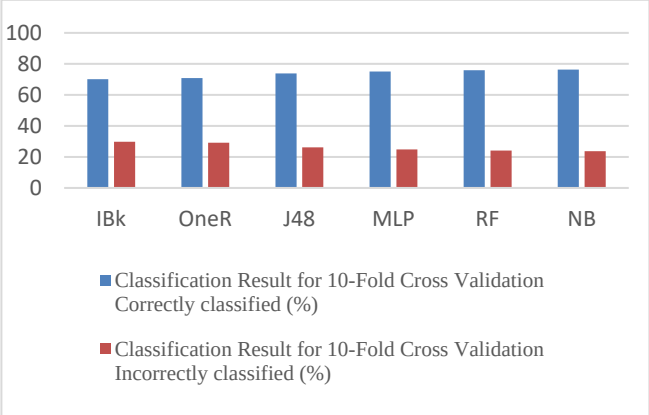

Fig. 12 Classifiers using 10-Fold Cross Validation

Table 7 and Table 8 show different data mining results of classifier precision, recall and F-Measures for NaiveBayes (NB), MultiLayer Perceptron (MLP), OneR, J48 IBk, J48, OneR, MultiLayer Perceptron (MLP), RandomForest (RF), MultiLayer Perceptron (MLP), and NaiveBayes (NB) algorithms using Weka for 5-fold and 10-fold cross-validation measures. NaiveBayes produces the highest value of 67.8% precision, 61.2% recall and 64.3% F-measure among other available algorithms in 5-fold cross-validation and 67.6% precision, 61.6% recall and 64.5% F-measure among other available algorithms in 10-fold cross-validation.

Table 7. Precision, Recall and F-Measure Performance of Classifier - 5-Fold Cross Validation

Name of the Classifiers	Precision	Recall	F-Measure
OneR	62.4	48.9	54.8
IBk	58.1	53.4	55.6
J48	58.5	60.4	59.4
Multilayer perceptron	64	60.4	62.2
RandomForest	66.3	63.1	64.6
NaiveBayes	67.8	61.2	64.3

Table 8. Precision, Recall and F-Measure Performance of Classifier-10-Fold Cross Validation

Name of the Classifiers	Precision	Recall	F-Measure
OneR	60.5	47.4	53.1
IBk	58	53	55
J48	63.2	59.7	61.4
MLP	65.3	61.2	63.2
RF	66.7	61.9	64.2
NB	67.6	61.6	64.5

The classifiers' error rate is demonstrated in Table 9 and Table 10 for 5-fold cross-validation and 10-fold cross-validation, respectively. The MAE of NB in 5-fold cross-validation and 10-fold cross-validation, a value of 28.43, is obtained. Hence, the error rate of NB is highly reduced compared with all other available algorithms.

Table 9. Classifier Error Performance using 5-fold cross-validation

Classifiers in %	One R	J-48	RF	IBk	MLP	NB
Mean Absolute Error (MAE)	34.83	32	31.23	29.7	29.36	28.43
Root Mean Squard Error (RMSE)	28.13	46	40.45	54.4	41.58	41.69
Relative Absolute Error (RAE) (%)	53.03	71	68.70	65.4	64.58	62.47

Table 10. Classifier Error Performance using 10-fold cross-validation

Classifiers in %	J-48	IBk	OneR	MLP	RF	NB
MAE	32.19	29.88	29.12	29.38	31.29	28.43
RMSE	46.45	54.53	54.01	42.36	40.52	41.69
RAE	70.81	65.73	64.17	64.64	68.84	62.54

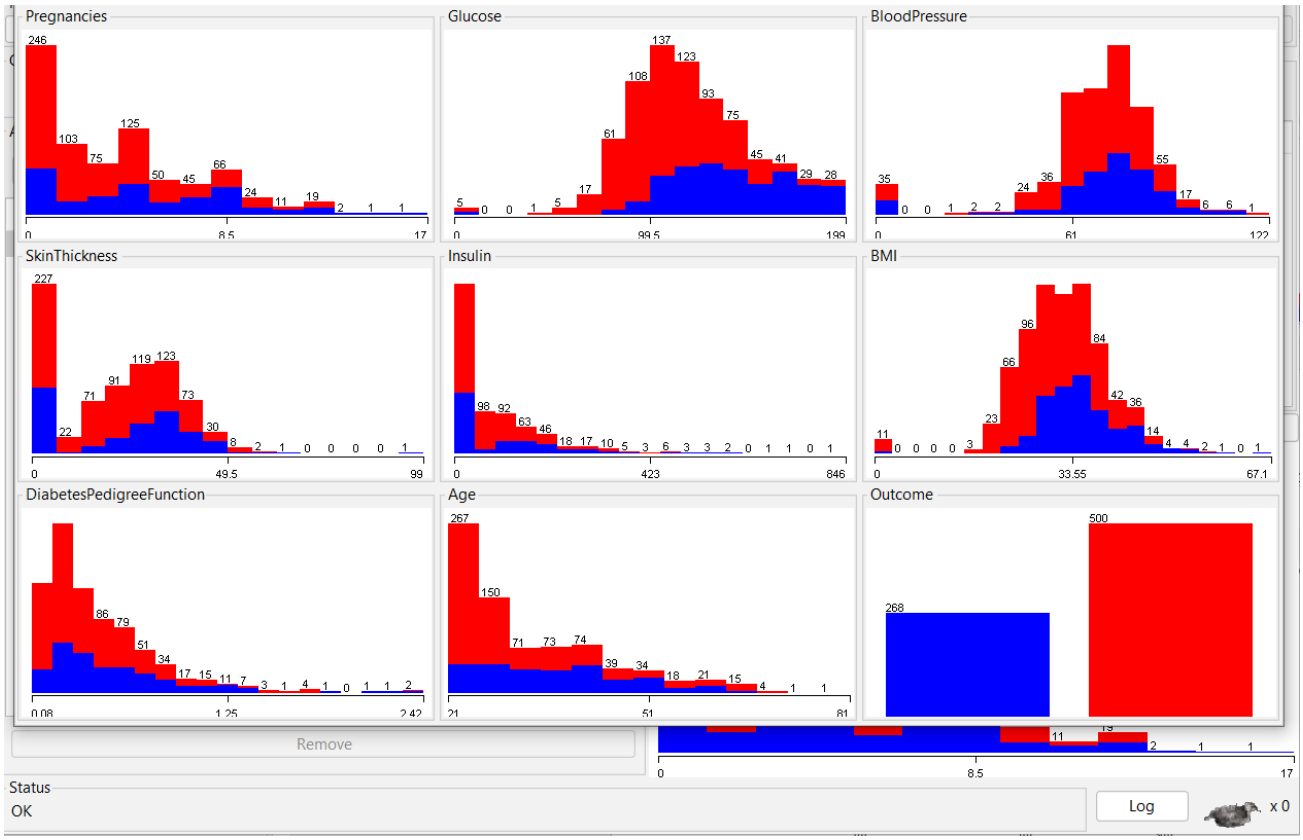

Fig. 13 Visualization of PID dataset features

```
14/14 [==============================] - 0s 5ms/step - loss: 0.4093 - accuracy: 0.8032 - val_loss: 0.4659 - val_accuracy: 0.7778
Epoch 193/200
14/14 [==============================] - 0s 7ms/step - loss: 0.4184 - accuracy: 0.8056 - val_loss: 0.4659 - val_accuracy: 0.7778
Epoch 194/200
14/14 [==============================] - 0s 5ms/step - loss: 0.4296 - accuracy: 0.8032 - val_loss: 0.4662 - val_accuracy: 0.7778
Epoch 195/200
14/14 [==============================] - 0s 6ms/step - loss: 0.3964 - accuracy: 0.8079 - val_loss: 0.4664 - val_accuracy: 0.7778
Epoch 196/200
14/14 [==============================] - 0s 6ms/step - loss: 0.4071 - accuracy: 0.8241 - val_loss: 0.4669 - val_accuracy: 0.7778
Epoch 197/200
14/14 [==============================] - 0s 5ms/step - loss: 0.4055 - accuracy: 0.8056 - val_loss: 0.4665 - val_accuracy: 0.7778
Epoch 198/200
14/14 [==============================] - 0s 5ms/step - loss: 0.4158 - accuracy: 0.7894 - val_loss: 0.4675 - val_accuracy: 0.7708
Epoch 199/200
14/14 [==============================] - 0s 5ms/step - loss: 0.4070 - accuracy: 0.8194 - val_loss: 0.4673 - val_accuracy: 0.7778
Epoch 200/200
14/14 [==============================] - 0s 5ms/step - loss: 0.4317 - accuracy: 0.7986 - val_loss: 0.4666 - val_accuracy: 0.7778
```

Fig. 14 Training and Testing Accuracy of IANN

Visualization of data is an essential part of data science. It supports knowing data and can be used to explain the data to a new person. The visualization of all features in the PID dataset is shown in Figure 13.

IANN evaluated the model with training and validation accuracy. It is represented in Figure 14 with an epoch of 200. This deep learning model produces a training accuracy of 79% and a testing accuracy of 77%. Hence, this model produces the highest accuracy compared to existing algorithms. Figure 15 shows the result of the proposed classifier, which produces 83% precision, 81% recall, and 82% F-Measure. The proposed IANN produces good precision and F-measure compared to the existing NB. Figure 16 exploded the proposed IANN 10-fold cross-validation results. The 10th fold gets 69.74% accuracy, and the average accuracy is 66%.

```
print(classification_report(results,y_test))
```

	precision	recall	f1-score	support
0	0.83	0.81	0.82	126
1	0.65	0.68	0.67	66
accuracy			0.77	192
macro avg	0.74	0.75	0.74	192
weighted avg	0.77	0.77	0.77	192

Fig. 15 Proposed work-classifier results.

```
Epoch 13/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6593 - accuracy: 0.6532
Epoch 14/20
22/22 [==============================] - 0s 3ms/step - loss: 0.6553 - accuracy: 0.6532
Epoch 15/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6566 - accuracy: 0.6517
Epoch 16/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6539 - accuracy: 0.6532
Epoch 17/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6565 - accuracy: 0.6474
Epoch 18/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6528 - accuracy: 0.6532
Epoch 19/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6516 - accuracy: 0.6532
Epoch 20/20
22/22 [==============================] - 0s 2ms/step - loss: 0.6515 - accuracy: 0.6532
accuracy: 69.74%
65.50% (+/- 1.65%)
```

Fig. 16 Result of 10 fold cross validation

5. Conclusion and Future Direction

This research uses the PID dataset to build the deep learning model for diabetic classification. The Weka tool is used to build the classifiers for IBk, OneR, J48, Multilayer Perceptron, RandowForest, and Naive Bayes. The Nave Bayes model produces good performance compared to other existing models. Our proposed improved artificial neural network model is designed as a modified ANN using multi-layer perceptron, producing 79% training accuracy and 77% testing accuracy. Furthermore, this work can be extended using a huge dataset and deep learning techniques.

References

[1] Brenda Robles, and Tony Kuo, "Provider–Patient Interactions as Predictors of Lifestyle Behaviours Related to the Prevention and Management of Diabetes," *Dialectology*, vol. 3, no. 1, pp. 176-192, 2022. *Crossref*, https://doi.org/10.3390/Diabetology3010012

[2] Rajput SA, Ashraf S, and Siddiqui M, "Diet and Management of Type II Diabetes Mellitus in the United Kingdom: A Narrative Review," *Diabetology*, vol. 3, no. 1, pp. 72-78, 2022. *Crossref*, https://doi.org/10.3390/Diabetology3010006

[3] Chayal Angermanangus Forbesglenn Robert, "The Experiences of Insulin Use Among Older People with Type 2 Diabetes Mellitus: A Thematic Synthesis," *Primary Care Diabetes*, vol. 16, no. 5, pp. 614-626, 2022. *Crossref*, https://doi.org/10.1016/J.Pcd.2022.08.008

[4] Ehab Mudher Mikhael et al., "Self-Management Knowledge and Practice of Type 2 Diabetes Mellitus Patients in Baghdad, in: Diabetes Metabsyndrobes, a Qualitative Study," *Diabetes, Metabolic Syndrome and Obesity:Targets and Therapy*, vol. 12, pp. 1–17, 2019.

[5] E.M. Bayked, B.D. Workneh, and M.H. Kahissay, "Thirst for Information and Needs Reflections of Type 2 Diabetes Patients Receiving Insulin Treatment in North-East Ethiopia: A Qualitative Exploration," *Journal of Preventive Medicine and Public Health*, vol. 54, no. 2, pp. 119–128, 2021.

[6] D. Sheema et al., "An Algorithm for Detection and Identification of Infestation Density of Pest-Fall Armyworm in Maize Plants Using Deep Learning Based on IoT," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 240-251, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I9P224

[7] Alaa KH.Mashni, Dr.Tamer Issawi, and Dr.Muhammad Farraj, "Molecular Characterization of Type 2 Diabetes Mellitus By Single Nucleotide Polymorphism of Transcription Factor 7 Like 2 Gene," *SSRG International Journal of Medical Science*, vol. 5, no. 9, pp. 1-9, 2018. *Crossref*, https://doi.org/10.14445/23939117/IJMS-V5I9P101

[8] Jenifer Florence Mary J, and amol R.Dongre, "Effect of a Community-Based Intervention on Self-Care Among Diabetes Patients in Rural Tamil Nadu: A Mixed-Method Study," *Primary Care Diabetes*, vol. 16, no. 4, pp. 484-490, 2022. *Crossref*, https://doi.org/10.1016/J.Pcd.2022.05.001.

[9] Xiangli et al., "Effect of Glucagon-Like Peptide 1 Receptor Agonists on the Renal Protection in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis," *Diabetes & Metabolism*, vol. 48, no. 5, P. 101366, 2022. *Crossref*, https://doi.org/10.1016/J.Diabet.2022.101366.

[10] Shakti Chourasiya, and Suvrat Jain, "A Study Review on Supervised Machine Learning Algorithms," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 8, pp. 16-20, 2019. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V6I8P104

[11] Olajide Tayo Emmanuel et al., "Assessment of Knowledge About Diabetes Mellitus and Self-Management Among Patients Attending Two Teaching Hospitals in Ogun State," *SSRG International Journal of Medical Science*, vol. 6, no. 7, pp. 22-27, 2019. *Crossref*, https://doi.org/10.14445/23939117/IJMS-V6I7P105

[12] Robyn J Tapp et al., "Longitudinal Association of Glucose Metabolism with Retinopathy: Results From the Australian Diabetes Obesity and Lifestyle (Ausdiab) Study," *Diabetes Care*, vol. 31, pp. 1349-1354, 2008.

[13] Miyazaki M et al., "Comparison of Diagnostic Methods for Diabetes Mellitus Based on Prevalence of Retinopathy in a Japanese Population: the Hisayama Study," *Diabetologia*, vol. 47, no. 8, pp. 1411-1415, 2004. *Crossref*, https://doi.org/10.1007/s00125-004-1466-8

[14] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care*, vol. 33, pp. S62-S69, 2010.

[15] Dr. R. Surendiran et al., "Effective Autism Spectrum Disorder Prediction to Improve the Clinical Traits Using Machine Learning Techniques," *International Journal of Engineering Trends and Technology*, vol. 70, no. 4, pp. 343-359, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I4P230

[16] Diagnosis and Classification of Diabetes Mellitus, *Diabetes Care*, vol. 33, pp. S62-S69, 2010.

[17] Sankalp Yadav, and Gautam Rawal, "Role of Integrating Community Health Workers in Achieving Healthcare Information for All," *International Journal of Scientific Research and Reviews*, vol. 4, no. 1, pp. 106-10, 2015.

[18] Sankalp Yadav, and Gautam Rawal, "Counterfeit Drugs: Problem of Developing and Developed Countries," *International Journal of Pharmaceutical Chemistry and Analysis*, vol. 2, no. 1, pp. 46-50, 2015.

[19] Sankalp Yadav, Gautam Rawal, and Mudit Baxi, "Zika Virus- A Pandemicin Progress," *Journal of Translational Internal Medicine*, vol. 4, no. 1, pp. 42-45, 2016. *Crossref*, https://doi.org/10.1515/jtim-2016-0009

[20] Aurélie Berot et al., "Characteristics of Newly Diagnosed Type 1 Diabetes in Pediatrics and Adult Population From Reims University Hospital, France From 1997 to 2019," *Diabetes & Metabolism*, vol.48, no. 5, p. 101346, 2022. *Crossref*, https://doi.org/10.1016/J.Diabet.2022.101346.

[21] Anda-Petronelaradan et al., "Gestational Diabetes is Associated with SARS-Cov-2 Infection Duringpregnancy: A Case-Control Study," *Diabetes & Metabolism*, vol. 48, no. 4, p. 101351, 2022. *Crossref*, https://doi.org/10.1016/J.Diabet.2022.101351.

[22] Dr. R. Surendiran et al., "A Systematic Review Using Machine Learning Algorithms for Predicting Preterm Birth," *International Journal of Engineering Trends and Technology*, vol. 70, no. 5, pp. 46-59, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I5P207

[23] Haider MB, Abbas F, and Hafeez W, "A 46-Year-Old Woman who Presented with Diabetic Ketoacidosis and COVID-19 Pneumonia with Multiple Pulmonary Thromboemboli: A Case Report," *The American Journal of Case Reports*, vol. 21, p. E925794, 2020. *Crossref*, https://doi.org/10.12659%2FAJCR.925794

[24] Kok Hoe Chan et al., "Clinical Characteristics and Outcome in Patients with Combined Diabetic Ketoacidosis and Hyperosmolar Hyperglycemic State Associated with COVID-19: A Retrospective, Hospital-Based Observational Case Series," *Diabetes Research and Clinical Practice*, vol. 166, p. 108279, 2020. *Crossref*, https://doi.org/10.1016%2Fj.diabres.2020.108279

[25] Francisco J Pasquel et al., "Clinical outcomes in Patients with Isolated or Combined Diabetic Ketoacidosis and Hyperosmolar Hyperglycemic State: A Retrospective, Hospital-Based Cohort Study," *Diabetes Care*, vol. 43, P. 349e57, 2020.

[26] Nina Goldman et al., "High Prevalence of COVID-19-Associated Diabetic Ketoacidosis in UK Secondary Care," *Diabetes Research and Clinical Practice*, vol. 166, p.108291, 2020. *Crossref*, https://doi.org/10.1016/j.diabres.2020.108291

[27] Dr. Surendiran R et al., "Exploring the Cervical Cancer Prediction by Machine Learning and Deep Learning with Artificial Intelligence Approaches," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 94-107, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I7P211

[28] Dipo Theophilus Akomolafe et al, "Resolving Some Critical Issues in the Prevention, Diagnosis, Treatment and Management of Covid-19 Using Machine Learning," *International Journal of Computer and Organization Trends*, vol. 10, no. 4 , pp. 1-8, 2020. *Crossref*, https://doi.org/10.14445/22492593/IJCOT-V10I4P301

[29] Aysha Asif Sarwani et al., "Clinical Outcomes of Diabetic Ketoacidosis in Adult Patients in Bahrain," *SSRG International Journal of Medical Science*, vol. 8, no. 2, pp. 1-9, 2021. *Crossref*, https://doi.org/10.14445/23939117/IJMS-V8I2P101

[30] D.J. Newman et al., "*UCI Repository of Machine Learning Databases*," Department of Information and Computer Science, University California Irvine, 1998.

[31] Meijian Liao et al., "A Group of Long Noncoding RNAs Identified by Data Mining Can Predict the Prognosis of Lung Adenocarcinoma," *Cancer Science*, vol. 109, no. 12, pp. 4033-4044, 2018. *Crossref*, https://doi.org/10.1111/cas.13822

[32] Pawar S, "Supervised Machine Learning Algorithms for Accurately Classifying Cancer Types from Complex Gene Expression Systems," *In IX International Conference on Complex Systems*, p. 202, 2018.

[33] S. Bhuvaneswari et al., "Disease Detection in Plant Leaf Using LNET Based on Deep Learning," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 64-75, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I9P207

[34] Zulaiha Parveen A, and Mr.T.Senthil Kumar, "The Deep Learning Methodology for Improved Breast Cancer Diagnosis in MRI," *International Journal of Computer and Organization Trends*, vol. 11, no. 3, pp. 11-14, 2021. *Crossref*, https://doi.org/10.14445/22492593/IJCOT-V11I3P303

[35] Meteb Altaf et al., "Monitoring Diabetic Foot by Designing a New Smart Sole," *SSRG International Journal of Pharmacy and Biomedical Engineering*, vol. 7, no. 3, pp. 15-20, 2020. *Crossref*, https://doi.org/10.14445/23942576/IJPBE-V7I3P103

[36] Xiaoyan Lv et al., "Self-Care Experiences of Empty-Nest Elderly Living with Type 2 Diabetes Mellitus: A Qualitative Study from China," *Frontiers in endocrinology*, vol. 12, p. 745145, 2021. *Crossref*, https://doi.org/10.3389/fendo.2021.745145

[37] Dr Olutoyin Elizabeth Okeya, "Evaluating Self-Management Experience of Type 2 Diabetes Among Black Africans in Liverpool," *SSRG International Journal of Medical Science*, vol. 7, no. 4, pp. 5-11, 2020. *Crossref*, https://doi.org/10.14445/23939117/IJMS-V7I4P102

[38] *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation*, vol. 93, pp. 299-309, 2011.

[39] Rimesh Pal et al., "Clinical Profile and Outcomes in COVID-19 Patients with Diabetic Ketoacidosis: A Systematic Review of Literature," *Diabetes & Metabolic Syndrome*, vol. 14, no. 6, pp. 1563-1569, 2020. *Crossref*, https://doi.org/10.1016/J.Dsx.2020.08.015

[40] Yash Panchori et al., "Malaria Parasite Detection Using Various Machine Learning Algorithms and Image Processing," *SSRG International Journal of Computer Science and Engineering* , vol. 7, no. 2, pp. 68-70, 2020. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V7I2P108

[41] Emily Jane Gallagher, Derek Le Roith, and Zachary Bloomgarden, "Review of Hemoglobina1cin Themanagement of Diabetes," *Journal of Diabetes*, vol. 1, pp. 9-17, 2009. *Crossref*, https://doi.org/10.1111/j.1753-0407.2009.00009.x

[42] Gautamrawal et al., "Glycosylated Hemoglobin (Hba1c): A Brief Overview for Clinicians," *Indian Journal of Immunology and Respiratory Medicine*, vol. 1, no. 2, pp. 33-36, 2016.

[43] S. Willner, R. Whittemore, and D. Keene, "Life or Death, Experiences of Insulin Insecurity Among Adults with Type 1 Diabetes in the United States," *SSM - Population Health*, vol. 11, p. 100624, 2020. *Crossref*, https://doi.org/10.1016/j.ssmph.2020.100624