International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Prediction of Diabetes Mellitus Using Machine learning and Deep Learning models

Afshan Fatima[1], Saurabh Pal[2], Venkateswara Rao Ch[3]
[1]Research Scholar, CSE Department, Vbs Purvanchal University
[2]Professor, CSE Department, Vbs Purvanchal University
[3]Associate Professor, Department of CSE, Siddhartha Institute of Engineering& Technology
[1]afshafatima0889@gmail.com
[2]drsaurabhpalvbspu@gmail.com
[3]chvenkatsrh@gmail.com

Abstract: In this contemporary era, the global community finds itself profoundly impacted by a prevalent endocrine disorder known as Diabetes Mellitus, more commonly referred to as diabetes. With the ever-increasing urgency for an efficacious framework that can prognosticate diabetes and its various classifications at the nascent stages, a novel and cutting-edge Machine Learning based Model (MLM) has been put forth as a solution. This approach MLM possesses the unique capability to accurately forecast the occurrence of diabetes, as well as categorize it into distinct types, namely type 1, type 2, and Gestational diabetes, afflicting the patient. The advent of this pioneering MLM marks a significant milestone in the realm of diabetes diagnosis, surpassing the precision of its predecessors and presenting a fusion of expertise harnessed from both expert systems and the machine learning milieu.

Keywords: Artificial Intelligence, Machine Learning, Machine Learning based Model, Data Mining, Diabetes Mellitus, Classification, Diabetes, Healthcare.

1. INTRODUCTION

The “Artificial intelligence (AI)” is poised to revolutionize the healthcare sector, making a profound impact on various facets of the industry. The application of data science and machine learning in healthcare can provide invaluable insights for analysis and prediction. With its potential in medical image analysis, drug discovery, genetics research, and predictive medicine, healthcare stands as a promising domain for the implementation of data science solutions [1].

As this technology becomes more prevalent in the healthcare field, there is an expectation that AI will greatly assist in the prognosis and diagnosis of a wide range of diseases. This ability to emulate human decision-making expertise holds tremendous potential. To put things into perspective, the “International Diabetes Federation (IDF)” estimates that there are approximately 415 million people living with diabetes worldwide, with India being the second-largest country with a staggering 69 million diabetics [2].

These numbers serve as a catalyst for the ongoing research in this area, highlighting the need for a computerized system that can accurately predict the onset of diabetes in its early stages. Diabetes, a disorder of carbohydrate metabolism, is characterized by the body's impaired ability to produce or respond to insulin, resulting in improper glucose levels in the blood. It is broadly classified into three main types: “type 1 diabetes, type 2 diabetes, and gestational diabetes” [3].

The “Type 1 diabetes”, which constitutes only 5-10% of the total diabetes population, occurs due to the destruction of the islets of Langerhans in the pancreas, often as a result of autoimmune dysfunction. These islets are responsible for insulin production. On the other hand, “type 2 diabetes”, accounting for 90-95% of the total diabetes population, is closely linked to obesity and is characterized by both insulin resistance and deficiency.

Lastly, “gestational diabetes” is a temporary condition that occurs during pregnancy, leading to elevated blood glucose levels. Fortunately, this condition usually resolves itself after delivery. However, it is important to note that women who have experienced gestational diabetes are at an increased risk of developing diabetes in the future.

The authors of this paper have put forth an innovative approach to anticipate the occurrence of diabetes and its various manifestations. This groundbreaking predictive research endeavors to empower individuals by providing them with insights into their personal inclination to diabetes, thereby enabling them to take proactive measures to tackle this prevalent healthcare concern. The fundamental objective of this study is to construct a sophisticated “Machine Learning Model (MLM)” that can effectively diagnose diabetes. The accurate identification of pertinent symptoms exhibited by patients plays a pivotal role in the successful application of diagnostic rules. These amalgamations of knowledge serve as determining factors in discerning whether an individual is afflicted with diabetes or not, if yes then its type.

A specific assessment of the MLM was conducted on a group of 10 patients, yielding results that align precisely with those of medical professionals. The MLM devised by the researchers possesses the potential to be employed with utmost efficiency in diagnosing different types of diabetes. Notably, this intelligent MLM promises to be particularly valuable in regions where the scarcity of medical practitioners poses a significant challenge. By diminishing reliance on healthcare providers, this pioneering technology aims to enhance diagnostic accuracy and expediency for both medical professionals and patients [2].

2. Literature Review

Extensive research has been conducted in recent years, employing various “machine learning (ML) and deep learning (DL)” algorithms, to forecast the onset of diabetes mellitus [1]. However, the majority of these investigations have primarily focused on refining the predictive model itself. Leveraging the TensorFlow platform, the Sequential function was harnessed to construct the diabetes prediction model [2]. Subsequently, the model was converted into the 'tflite' format, facilitating its integration into the mobile application through the Android Studio integrated development environment (IDE) [3].

The development of this pioneering deep learning-powered mobile application represents a significant technological breakthrough in the early detection of diabetes mellitus [4]. By enabling timely predictions, individuals can proactively modify their lifestyle choices, thereby averting potentially grave complications [5].

Undetected and untreated diabetes can give rise to a plethora of health complications, including “cardiovascular afflictions and cerebrovascular incidents” [6]. A model that possesses the capability to prognosticate the likelihood of an individual succumbing to diabetes with a commendable degree of accuracy [7]. In order to identify diabetes in its nascent stages, an assemblage of six supervised ML classification techniques, along with a hybrid model predicated upon the uppermost triumvirate of discoveries, is employed [8]. The repository of ML data proffered by UCI, known as the “Pima Indians Diabetes Database”, serves as the foundation, which exhibits a discerning accuracy rate of 90.62%, through highbred approach [9].

The “Diabetes Mellitus” stands as a metabolic chronic ailment in which the levels of glucose in the bloodstream surpass the normal range. Remarkably, a staggering 8.7% of the Indian populace, spanning the age bracket of 20 to 70, grapple with this condition [10]. The repercussions of undetected and untreated diabetes extend beyond measure, encompassing detrimental impacts on vital organs such as the heart, kidneys, eyes, nerves, and blood vessels [11]. While existing clinical decision-making methods strive valiantly to address the multifarious challenges presented by this malady, there remains a pressing need for further enhancements [12].

The process involves that the dataset undergoes meticulous preprocessing, wherein the normalization technique takes center stage, ensuring optimal data integrity through "deep belief neural network" assumes its rightful place as the harbinger of prognostic precision including the likes of "naïve Bayes, Decision Tree, Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM)" [13].

In the year 2019, approximately 9.3 percent of adults worldwide were afflicted by "Diabetes Mellitus", the projected figure for the year 2045 indicates a staggering increase, affecting nearly 11 percent of the global population [14]. To anticipate the onset of diabetes Mellitus may result in severe damage to vital organs such as nerves, eyes, and kidneys along with the timely prediction and subsequent treatment of this condition hold the potential to save countless lives through DL models that exhibit unparalleled accuracy in predicting diabetes when contrasted with alternative models using the "PIMA Indians Diabetes database" for diagnosing [15-20].

In recent times, computerized physician consultants have gained significant popularity as invaluable tools for enhancing the overall health status of individuals and these systems necessitate extensive research into emerging diseases, ultimately contributing to the development of clinical decision support systems [21]. A web-based application has been created to predict the risk of early-stage diabetes in patients as the proposed application utilizes the classification outcomes derived from patient questionnaires, effectively eliminating the need for traditional testing kits [22]. These predictions are further analyzed and categorized into distinct clusters using unsupervised machine learning algorithms, enabling the identification of positive or negative diabetes indicators [23].

3. PROPOSED METHODOLOGY

3.1 Dataset Used:

This dataset encompasses complex details pertaining to symptoms and classifications of diabetes types, correlated with specific values assigned to each symptom. A comprehensive total of 64 symptoms have been precisely considered to categorize diabetes into three distinct types: "Type-1 Diabetes, Type-2 Diabetes, and Gestational Diabetes". Consequently, the dataset comprises a sum of 65 columns, with 64 denoting symptoms and the final column representing the prediction value termed as "PredictionOfDiabetes". The dataset is prepared in such a manner that each row therein symbolizes a singular test case for each diabetes type. These rows serve as the foundation for the ML model with a set of rules aptly labeled as a "knowledge base". To be precise, a total of 120 records have been meticulously entered into the dataset, essentially signifying the complete set of rules governing the machine learning model. It is crucial to underscore that the process of designing these rules and selecting the optimal rule for the machine learning model is an intricate endeavor, often necessitating the expertise of knowledgeable individuals to enhance the accuracy of predictions.

3.2 Designing the Knowledge Base for Machine Learning Model:

The dataset encompasses predominant indications the diabetes symptoms which have been systematically categorized based on their respective types, and the tabular format has been meticulously formulated utilizing numerical values. These values, which assume a

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

binary nature, manifest as either zero or one, signifying the presence or absence of a symptom. However, in the case of symptoms that possess a binary nature, wherein they can be classified as either true or false, the corresponding binary values assigned to them are zero and one, as illustrated in Table 2. To elucidate further, the binary value of zero represents the absence of a Family History, while a binary value of one represents its presence. Conversely, symptoms that exhibit a multi-valued nature are assigned a series of numbers. For instance, the Age category is demarcated into three distinct groups: Young, Adult, and Old, which are represented by the numerical values 11, 12, and 13, respectively. Similarly, Obesity is stratified into three categories: Low, Normal, and Obese, which are denoted by the numerical values 21, 22, and 23, correspondingly. Likewise, Hypertension encompasses four classifications: Normal, Elevated, High, and Very High, which are connoted by the numerical values 3, 4, 5, and 6, respectively. HDL Cholesterol demonstrates a trichotomous division: Low, Medium, and High, indicated by the numerical values 1, 2, and 3, respectively. Triglyceride, on the other hand, encompasses four classifications: Normal, Borderline, High, and Very High, represented by the numerical values 2, 3, 4, and 5, respectively. It is important to note that Table 1 has been meticulously curated through an amalgamation of primary data collection, consisting of information obtained from physicians, medical literature, online resources, and medical journals [4], as well as data derived from diabetic patients. This meticulously compiled table serves as one of the three pivotal inputs for the system. Moreover, while designing rules in the CSV file, a varied combination of symptoms for each type of diabetes has been taken into consideration.

Table 1:Dataset as a Knowledge base

Sl.No.	Symptoms	A	B	CS	Sl.No.	Symptoms	A	B	CS
1	family_History	0	1	0	33	Vomiting	1	0	0
2	Age	11	12	12	34	Dry_Skin	0	1	0
3	Obesity	22	23	21	35	Aches&Pains	0	1	0
4	evious_IFG_IGT	0	1	0	36	Recurresnt_fungal_infectn	0	1	0
5	Hypertension	0	5	5	37	Nightmares	1	1	0
6	HDL_Cholesterol	0	3	0	38	Seizures	1	1	0
7	igyceride	0	5	0	39	Sadness	1	1	0
8	dceased_Thirst	1	1	0	40	Unconsciousness	1	1	0
9	Icreased_Urinate	1	1	0	41	Numbness	1	1	0
10	Increased_Appetite	1	1	0	42	VaginalMycoticInfectn	1	1	1
11	Weight_Variation	0	1	0	43	Rapid_Heart_Beat	0	1	0
12	Impaired_Vision	0	1	0	44	Recurring_Gum_Infe	0	0	0
13	Tiredness	1	1	0	45	Impotency	1	1	0
14	Impatience	0	1	0	46	high_bloodPressure	0	0	1
15	Infection	0	1	0	47	Sleep_Walking	1	1	0
16	Itchy_Skin	0	0	0	48	Makeg_unusual_noises	1	1	0
17	Depression_Stress	0	0	0	49	Leg_Cramps	1	1	0
18	Tingling_Sensation	0	0	0	50	Slurred_Speech	1	1	0
19	Fruity_Breath_Odour	1	0	0	51	Flushed_face	1	1	0
20	Bed_Wetting	1	0	0	52	Pale_Skin	1	0	0
21	Slow_Healing_Wound	0	1	0	53	LossOfMenstruation	1	1	0
22	FamilyHis_Pregnancy	0	0	1	54	Stomach_Pain	1	1	0
23	Previous_Pregnancy	0	0	1	55	Deep_Breathing	1	1	0
24	BabyOver_9Pd_PrePreg	0	0	1	56	Areas_Darked_Skin	0	1	0
25	Sleeplessness	1	1	0	57	Difficult_Concentrating	1	1	0
26	Trembling	1	1	0	58	Dehydration	1	1	0
27	Sweating	1	1	0	59	LackOfCoordination	1	1	0
28	Anxiety	1	1	0	60	Hist_Heart_Stroke	0	0	0
29	Confusion	1	1	0	61	Poly_Ovary_Syndrome	0	0	0
30	Weakness	1	0	0	62	bloodSugar_NewbornBabqy	0	0	1
31	Mood_Swings	1	0	0	63	WaistSize02cmM88cmF	0	0	0
32	Nausea	1	0	0	64	WaistHipRatio.9M.85F	0	0	0

Table 2: Number representation for symptoms

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Boolean Value Symptoms			Multi Value Symptoms		
Family History	*True*	1	**Age**	*Young*	11
	False	0		*Adult*	12
				Old	13
Obesity	*Low*	21	**Trigyceride**	*Normal*	2
	Normal	22		*Boarder Line*	3
	Obese	23		*High*	4
Hyper-tension	*Normal*	3		*Very High*	5
	Elevatal	4	**HDL_ Cholesterol**	*Low*	1
	High	5		*Medium*	2
	Very High	6		*High*	3

3.3 Implementation

The acquisition of secondary data comprises of a comprehensive exploration of various sources such as literature, online platforms, medical publications, and primary data collection through the dissemination of questionnaires to physicians, with the aim of discerning the symptoms that impact diabetes. The formulation of rules for this particular MLM was meticulously crafted, drawing upon thorough data analysis and expert recommendations. These rules were subsequently inputted into a CSV file, serving as the dataset for the machine learning model and facilitating subsequent processing. To facilitate the execution of the machine learning algorithm for this very model, the cloud-based service known as Google Colaboratory, or Colab in short, was harnessed, leveraging the power of Jupyter Notebooks. Within this dataset, one can uncover a wealth of knowledge pertaining to the MLM, manifesting as a series of rows peppered with numerical representations that encapsulate the symptoms associated with diabetes. The primary objective of this distinguished MLM is to predict the specific type of diabetes afflicting patients. Thus, the dataset assumes the role of a repository, housing both the rules and the knowledge base that fuel the MLM . Notably, within this rich repository of rules, one can discern the inclusion of three distinct categories of diabetes diagnoses, each of which can be interpreted as follows.

Family_History 1,Age 11, Obesity 22, Previous_IFG_IGT 0, Hypertension 0, HDL_Cholesterol 0 ,Trigyceride 0, Inceased_Thirst 1,Increased_Urinate 1, Increased_Appetite 1, Weight_Variation 0, Impaired_Vision 0, Tiredness 1, Impatience 0, Infection 0, Itchy_Skin 0, Depression_Stress 0 , Tingling_Sensation 0, Fruity_Breath_Odour 1, Bed_Wetting 1, Slow_Healing_Wound 0, FamilyHis_Pregnancy 0, Previous_Pregnancy 0, BabyOver_9Pd_PrePreg 0, Sleeplessness 1, Trembling 1, Sweating 1,Anxiety 1, Confusion 1, Weakness 1, Mood_Swings 1, Nausea 1,Vomiting 1, Dry_Skin 0, Aches&Pains 0, Recurresnt_fungal_infectn 0, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VaginalMycoticInfectn 1, Rapid_Heart_Beatm 0, Recurring_Gum_Infe 0, Impotency 1, high blood Pressure 0, Sleep_Walking 1, Makeg_unusual_noises 1, Leg_Cramps 1, Slurred_Speech 1, Flushed_face 1, Pale_Skin 1, LossOfMenstruation 1, Stomach_Pain 1, Deep_Breathing 1, Areas_Darked_Skin 0, Difficult_Concentrating 1, Dehydration 1, LackOfCoordination 1, Hist_Heart_Stroke 0, Poly_Ovary_Syndrome 0, LowbloodSugar_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0.Last Column as type 1 diabetes. Family_History 1,Age 12, Obesity 23, Previous_IFG_IGT 1, Hypertension 5, HDL_Cholesterol 3 ,Trigyceride 5, Inceased_Thirst 1,Increased_Urinate 1, Increased_Appetite 1, Weight_Variation 1, Impaired_Vision 1, Tiredness 1, Impatience 1, Infection 1, Itchy_Skin 0, Depression_Stress 0, Tingling_Sensation 0, Fruity_Breath_Odour 0, Bed_Wetting 0, Slow_Healing_Wound 1, FamilyHis_Pregnancy 0, Previous_Pregnancy 0, BabyOver_9Pd_PrePreg 0, Sleeplessness 1, Trembling 1, Sweating 1,Anxiety 1, Confusion 1, Weakness 0, Mood_Swings 0, Nausea 0,Vomiting 0, Dry_Skin 1, Aches&Pains 1, Recurresnt_fungal_infectn 1, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VaginalMycoticInfectn 1, Rapid_Heart_Beatm 1, Recurring_Gum_Infe 0, Impotency 1, high blood Pressure 0, Sleep_Walking 1, Makeg_unusual_noises 1, Leg_Cramps 1, Slurred_Speech 1, Flushed_face 1, Pale_Skin 0, LossOfMenstruation 1, Stomach_Pain 1, Deep_Breathing 1, Areas_Darked_Skin 1, Difficult_Concentrating 1, Dehydration 1, LackOfCoordination 1, Hist_Heart_Stroke 0, Poly_Ovary_Syndrome 0, LowbloodSugar_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0.Last Column as type 2 diabetes.

4. Machine Learning Based Model

The “decision tree classifier” serves as a systematic framework for engaging in multiclass classification endeavors. Its methodology involves posing a series of inquiries to the dataset at hand. Visualizing the “decision tree classification” algorithm reveals a binary tree structure, where each internal node and the root itself presents a distinct question. Consequently, the data residing within these nodes undergoes further division into individual records that possess unique characteristics. Ultimately, the leaves of the tree signify the distinct classes into which the dataset has been partitioned. To demonstrate the practical implementation of this approach, the following code snippet exemplifies how a researcher trains a decision tree classifier using scikit-learn. The “DecisionTreeClassifier” class within scikit-learn represents a powerful tool for conducting multi-class classification tasks with efficacy.

- **Importing Required Libraries and Dataset**: In this script, we begin by importing the necessary libraries for our task at hand. To handle our dataset in CSV format, the researcher utilizes the read_csv method from the pandas library (figure 1).
- **Obtaining the DataFrame from the dataset and storing the labels**: Next, we proceed with the data analysis phase. To gain insights into the structure of our dataset, the researcher employs the df.head() and df.shape commands. These commands provide information regarding the number of rows and columns present in our dataset.
- **Preparing the Data for Further Operations**: To facilitate further operations, we need to prepare our data appropriately. This involves dividing the data into attributes and labels. Subsequently, we further split the resulting data into training and test sets. This division enables us to train our algorithm effectively and subsequently test its performance.
- **Dividing the Data into Training and Test Sets**: To accomplish the aforementioned data division, we utilize the train_test_split method from the model_selection library of Scikit-Learn. This method randomly splits the data into training and testing sets.
- **Training the Decision Tree Algorithm and Making Predictions**: With our data now divided, we can proceed to train the decision tree algorithm on our training set and make predictions. The DecisionTreeClassifier class provides the predict method, which is employed for making predictions based on the trained algorithm.

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

- **Addressing the Issue of Overfitting**: It is important to note that overfitting can be a drawback when creating a knowledge base. However, this problem can be mitigated by utilizing suitable algorithms and exploring alternative solutions.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn

df = pd.read_csv("DiagnosisOfDiabetesTypesMLFinal.csv")
```

Figure 1. Importing libraries and Dataset

5. Results and Discussion

The proposed framework has been validated on a sample size of 10 individuals. Based on the patient's symptoms, the function "dtree_model.predict()" will take numerical parameters from Table 1 and Table 2. These symptoms serve as the foundation for the machine learning-based expert system to conduct predictions regarding the type of diabetes the patient may have. Five distinct prediction cases, validated for five patients, have been provided below. Subsequent to the collection of patient symptom data, this machine learning-based expert system generates the output, indicating the type of diabetes the patient is afflicted with.

Throughout the course of this research, an extensive evaluation of various models and expert systems for diagnosing diabetes was conducted in order to establish fundamental comparisons between the proposed system and existing ones. Onuiri Ernest E et al presented a web-based application that diagnoses diabetes by specifying symptoms and other relevant parameters, aiming to simulate an expert system. This application utilizes input parameters to determine the type of diabetes [5]. Conversely, Dilip Kumar Choubey et al developed a rule-based fuzzy expert system, which serves as a clinical decision support system for diabetes, expediting the diagnostic process. This rule-based expert system is capable of classifying diabetes into type 1, type 2, pre-diabetes, and gestational diabetes based on the presented symptoms. By adopting our proposed system, the probability of accurately diagnosing diabetes will be significantly enhanced [6]. Furthermore, our machine model incorporates additional symptoms, and the time and effort required for interface design will be reduced through the utilization of a machine learning platform.

We employed VP_Expert Primer to code the expert system for diabetes. Their decision table for diagnosis showcases various combinations of the patient's condition, age, symptoms, influential factors, and tests. By analyzing these factors, the system provides the final and we designed a mobile application to assess an individual's risk of developing diabetes, serving as an early prevention tool [8].

There exist numerous prognostic investigations encompassing the realm of social sciences [10], healthcare, and various other domains. The complexities associated with diabetes, such as diabetic neuropathy [11], retinopathy, and the like, also present extensive opportunities for the application of machine learning methodologies [12]. The realm of prediction is ceaseless, with an incessant demand for accurate and superior research. In order to attain heightened precision, ongoing predictive investigations are indispensable [13].

Consequently, future inquiries aim to identify more precise indicators while harnessing the potential of cutting-edge technologies to amplify the accuracy of prognostic rates to an extraordinary extent.

To predict and diagnose diabetes and its various types in patients utilizing machine learning models, researchers have ingeniously employed five distinct authentic case studies out of a total of ten patients. Astonishingly, these endeavors have yielded outcomes that align precisely with those of medical professionals.

Prediction Cases

Case1: the symptoms of the patient were passed as values from Table 1 and Table 2. The prediction was made using the dtree_model.predict() function, with the following symptoms: [0, 11, 22, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0].

```
[ ] My_Pred=dtree_model.predict([[0,11,22,0,0,0,0,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1

My_Pred

array(['Type1Diabetes'], dtype=object)
```

Figure 2. case 1 prediction as Type1Diabetes

The ML model processed these symptoms and generated a prediction for the type of diabetes the patient is suffering from. The output generated is: Type1Diabetes.

Case 2: Passing the corresponding symptom values for the second patient, as indicated in "Table 1 and Table 2", the model prediction for this patient (referred to as "My_Pred") is as follows: My_Pred = dtree_model.predict ([[1, 12, 23, 1,5, 3,5, 1,1, 1,1, 1,1, 1,1, 0,0, 0,0, 0,1, 0, 0,0,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,1,1,0,0,0,0,0]]).

```
[ ] My_Pred=dtree_model.predict([[1,12,23,1,5,3,5,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,0,0,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0

My_Pred

array(['Type2Diabetes'], dtype=object)
```

Figure 3. case 2 prediction as Type2Diabetes

The values provided are specific to the symptoms of the second patient and have been used to make a prediction using the dtree_model. And the output generated is: Type2Diabetes.

Case 3: Passing the symptom values for the third patient as indicated in Table 1 and Table 2. My_Pred=dtree_model.predict([[0,11,23,1,5,3,5,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,0,0,1,1,1,1,1,0,0, 0,0,1,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,1,1,0,0,0,0,0]]). And the output generated is: Type2Diabetes.

```
[ ] My_Pred=dtree_model.predict([[0,11,23,1,5,3,5,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,0,0,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0

My_Pred

array(['Type2Diabetes'], dtype=object)
```

Figure 4. case 3 prediction as Type2Diabetes

Case 4: The values of symptoms obtained from the fourth patient is referred from Table 1 and Table 2, are as follows: My_Pred = dtree_model.predict([[1,12,23,1,6,3,5,1,1,1,1,1,1,1,1,0,0, 0,0,0,1,0,0,0,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,1,1,1,0,0,0,0,0]]) and the output generated is: Type2Diabetes.

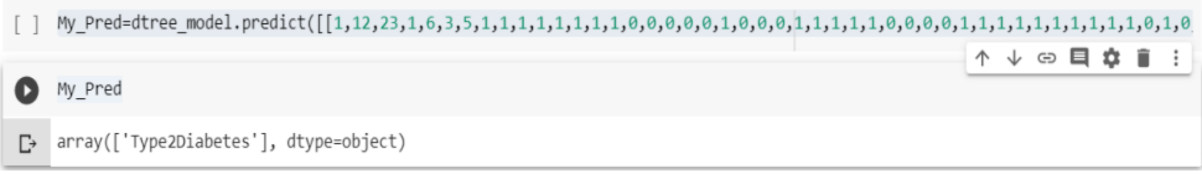

Figure 5. case 4 prediction as Type2Diabetes

Case 5: The values of symptoms obtained from the fifth patient, as referred from Table 1 and Table 2, are as follows: My_Pred = dtree_model.predict([[0,13,23,1,6,3,5,1,1,1,1,1,1,1,1,0, 0,0,0,0,1,0,0,0, 1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,1,1,1,0,0,0,0,0]]).

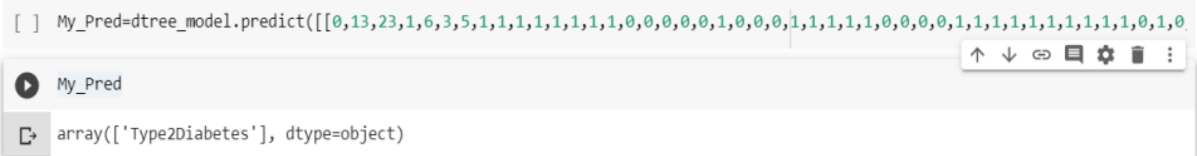

Figure 6. case 5 prediction as Type2Diabetes

And the output generated is: Type2Diabetes.

6. Conclusion

The proposed methodology in this paper involves the incorporation of machine learning competence with expert systems. By leveraging the non-parametric supervised learning technique known as the simple decision tree classification algorithm, we have devised a means to classify various types of diabetes with utmost precision. This model effectively anticipates the specific diabetes type based on discerning uncomplicated decision rules extracted from the salient data features. Consequently, this approach presents a comprehensive framework for seamlessly transforming expert systems into a formidable MLM platform. The perpetual quest for consistent and exceptional research in the realm of healthcare perpetually persists, as it engenders superior prognostic accuracy.

References:

[1] C. G. Estonilo and E. D. Festijo, "Development of Deep Learning-Based Mobile Application for Predicting Diabetes Mellitus," 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 2021, pp. 13-18, doi: 10.1109/IC2IE53219.2021.9649235.

[2] Mohammed Ali Shaik, “A Survey on Text Classification methods through Machine Learning Methods”, International Journal of Control and Automation (IJCA), ISSN:2005-4297, Volume12, Issue-6.

[3] S. Samet, M. R. Laouar and I. Bendib, "Diabetes mellitus early stage risk prediction using machine learning algorithms," 2021 International Conference on Networking and Advanced Systems (ICNAS), Annaba, Algeria, 2021, pp. 1-6, doi: 10.1109/ICNAS53565.2021.9628955.

[4] P. Prabhu and S. Selvabharathi, "Deep Belief Neural Network Model for Prediction of Diabetes Mellitus," 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), Singapore, 2019, pp. 138-142, doi: 10.1109/ICISPC.2019.8935838.

[5] Mohammed Ali Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, (2020), RNN based prediction of spatiotemporal data mining, 2020

IOP Conf. Ser.: Mater. Sci. Eng. 981 022027, doi.org/10.1088/1757-899X/981/2/022027
[6] E. Sabitha and M. Durgadevi, "Review on Diagnosis of Diabetes Mellitus Using Different Approaches," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 1-9, doi: 10.1109/ICECA52323.2021.9675879.
[7] C. Pankaj, K. V. Singh and K. R. Singh, "Artificial Intelligence enabled Web-Based Prediction of Diabetes using Machine Learning Approach," 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), Bengaluru, India, 2021, pp. 60-64, doi: 10.1109/CENTCON52345.2021.9688236.
[8] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, (2022), "COVID-19 Detector Using Deep Learning", International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 443-449, doi: 10.1109/ICAAIC53929.2022.9792694.
[9] E. Afzal, T. Saba, K. Javed, H. Ali and A. Karim, "Diagnosis and Prognosis of Diabetes Mellitus with Deep Learning," 2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU), Riyadh, Saudi Arabia, 2022, pp. 95-99, doi: 10.1109/WiDS-PSU54548.2022.00031.
[10] Whiting, David & Guariguata, Leonor & Weil, Clara & Shaw, Jonathan. (2011). IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes research and clinical practice. 94. 311-21. 10.1016/j.diabres.2011.10.029.
[11] Sherwani, S. I., Khan, H. A., Ekhzaimy, A., Masood, A. & Sakharkar, M. K. Significance of hba1c test in diagnosis and prognosis of diabetic patients. Biomarker Insights11, BMI–S38440 (2016).
[12] NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4*4 million participants. Lancet 2016; http://dx.doi.org/10.1016/S0140-6736(16)00618-8.
[13] Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: A global systematic review. PharmacoEconomics. 2015; 33(8): 811–31.
[14] Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041
[15] Nada Lavrac, "Selected techniques for data mining in medicine" , Artificial Intelligence in Medicine 16 (1999) 3–23
[16] Frawley W, Piatetsky-Shapiro G, Matheus C. Knowledge discovery in databases: an overview. In Piatetsky-Shapiro G, Frawley W, editors. Knowledge discovery in databases. Menlo Park, CA: The AAAI Press, 1991.
[17] HianChyeKoh and Gerald Tan,―Data Mining Applications in Healthcare, journal of Healthcare Information Management – Vol 19, No 2.
[18] Mohammed Ali Shaik and Dhanraj Verma, (2020), Enhanced ANN training model to smooth and time series forecast, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022038, doi.org/10.1088/1757-899X/981/2/022038
[19] Kincade, K. (1998). Data mining: digging for healthcare gold. Insurance & Technology, 23(2), IM2-IM7.
[20] Milley, A. (2000). Healthcare and data mining. Health Management Technology, 21(8), 44-47
[21] Christy, T. (1997). Analytical tools help health firms fight fraud. Insurance & Technology, 22(3), 22-26

[22] S.Yamini , Dr.V.Khanaa , Dr.Krishna Mohantha - A State of the Art Review on Various Data Mining Techniques, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 3, March 2016

[23] . Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238, Volume-29, Issue-4 (2020), Pp.169-175.