Abstract: Today world is extensively affected by endocrine disease Diabetes Mellitus which is commonly known as diabetes. There is a need for an effective model which can predict diabetes and its types at the early stages with accuracy. To improve the accuracy of prediction and to achieve better efficiency, a new Machine Learning based Model (MLM) is proposed. This Machine Learning Model (MLM) has ability to predict the diabetes and its categories as type 1, type 2 and Gestational diabetic with which the patient is suffering from. The proposed Machine Learning Model is innovative for diagnosis of diabetes is more accurate as compared to other existing approaches. This is a novel method from which one can combine power of an expert system with the machine learning environment.

Keywords: Artificial intelligent (AI), Machine Learning Model (MLM), Knowledge-base, Diabetes mellitus (DM), Decision Tree Algorithm.

I. INTRODUCTION

Artificial intelligence will have an outstanding impact on healthcare domain. Data science and machine learning can either be used for analysis or prediction in healthcare. Healthcare is a revolutionary and promising industry for implementing the data science solutions for Medical Images Analysis, Drug Discovery, Genetics Research and Predictive Medicine. As the technology becomes more common, in case of healthcare domain, it is expected that artificial intelligence will help prognosis, diagnose a wide area of disease. This emulates the decision making ability of a human expert. According to the International Diabetes Federation (IDF) it is estimated that 415 million people are living with diabetes in the world and India is the second largest country with 69 million diabetes people. These numbers are the motivation for the this respective research, where there is need of computerized system which will predict the diabetes in the early stages. Diabetes is a disorder of carbohydrate metabolism. This is characterized by an impaired ability of the body to produce insulin or respond insulin and thereby maintain proper levels of glucose (sugar) in the blood.

Diabetes broadly categorized into three main types as type 1 diabetes, type 2 diabetes, and gestational diabetes. In the case of type 1 diabetes, destruction of the islets of Langerhans of the pancreas happened because of autoimmune systems, this pancreas is source of insulin. From the total population of diabetes people it accounts only 5-10%. Type 2 diabetes is very common and accounts 90-95% of the total diabetes population, it’s strongly associated with obesity and is a result of resistance and deficiency of insulin. Gestational diabetes is a temporary condition in which blood glucose (sugar) levels increase during pregnancy state and it will return to normal conditions after delivery. The women, which are suffered from gestational diabetes has future risk of being diabetic[1].

In this research article authors have proposed a novel method for prediction of diabetes and its types. This predictive research will help an individual for knowing their own risk of being diabetic and the respective action should be done for avoiding such health care problem. The purpose of this study is to design a Machine Learning Model for diabetes diagnosis. Acquiring right symptoms of the patient is an important factor in the application of rules. These knowledge combinations determine whether a person is diabetes patient, along with its types such as type1 diabetes, type 2 diabetes and gestational diabetes. A Machine Learning Model was tested on 10 patients. It has achieved exact results as doctors. A Machine Learning Model that researcher have designed can be used effectively and efficiently to diagnoses for diabetes types. It helps patients in undeveloped countries where the number of doctors is not enough. This intelligent Machine Learning Model aims to reduce the dependence on doctors. It will help both doctors and patients to make more accurate and quicker decisions[2].

The lateral sections of this works are designed as follows II- Description of The Dataset, III- Designing The Knowledge Base for Machine Learning Model, IV- Designing The Knowledge Base For Machine Learning Model, V- Machine Learning Based Model VI. Results and Discussion VII. Conclusion and at last VII- References

II. DESCRIPTION OF THE DATASET

This dataset contains details of symptoms and diabetes type associated with the values given to each symptom. Total 64 symptoms are considered which will categorize diabetes into three different types as Type-1 Diabetes, Type-2 Diabetes, and Gestational Diabetes. Therefore dataset contains total 65 columns in which 64 are symptoms and last one with prediction value as PredictionOfDiabetes with 3 different values as Type-1 Diabetes, Type-2 Diabetes, and Gestational Diabetes. The researcher has designed the dataset such a way that each row in the dataset represents one test case for each diabetes type. These can be rule for the model and set of rules are termed as knowledge base. Therefore the dataset used for this Machine Learning Model itself is knowledge base for Machine Learning Model.

Total 120 records are entered into the dataset or on other hand those are total rule for the Machine Learning Model.

Revised Manuscript Received on January 15, 2020

* Correspondence Author
Manjiri Mahadev Mastoli*, Research Scholar, Department of Computer Science Shivaji University Kolhapur, Maharashtra, Email: it.manjiri@gmail.com
Dr.Urmila R. Pol, Assistant Professor, Department of Computer Science, Shivaji University Kolhapur, Maharashtra, India
Mr.Rahul D.Patil, Quality Assurance Engineer, Menon Bearing Ltd, Kolhapur, Maharashtra

Retrieved Number: E5916018520/202008EJESP
DOI:10.35940/ijrte.E5916.018520

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020
Designing the rules and selection of rule for Machine Learning Model is complicated task for which one can need expert knowledge which will enhance prediction accuracy.

III. DESIGNING THE KNOWLEDGE BASE FOR MACHINE LEARNING MODEL

The table 1 contains list of major symptoms of diabetic patients. Those are classified on the basis of their types and table has been designed with the help of values. These values are zero and one in the case symptoms which will take binary values and the series of numbers when symptoms are varies from more than two values. In case of symptoms which are having two categories either True OR False assigned binary values 0 and 1 (table 2). For example FamilyHistory represented by the binary value 0 and 1. In the case of multivalued symptoms it will take series of number. Age with three categories Young_Adult_Old represented by the number with values 11-12-13, Obesity with three categories Low_Normal_Obese represented by the number with values 21-22-23, Hypertension with four categories Normal_Elevatal_High_VeryHigh represented by the number with values 3-4-5-6, HDL_Cholesterol with three categories Low_Medium_High represented by the number with values 1-2-3, Trigyceride with four categories Normal_BoarderLine_High_VeryHigh represented by the number with values 2-3-4-5. This table 1 is prepared on the basis of primary data collection and secondary data collection such as physicians data, books, Internet, medical journals [4], diabetes patients etc. The Table 1 is considered as one of the three input for the system. In the same way with different combination of symptoms for each type of diabetes are considered while designing rule in CSV file.

Dataset as Knowledge base:

The table 1 contains list of major symptoms of diabetic patients. Those are classified on the basis of their types and table has been designed with the help of values. These values are zero and one in the case symptoms which will take binary values and the series of numbers when symptoms are varies from more than two values. Columns represented as A-Type1Diabetes, B- Type2Diabetes, C- Gestational Diabetes.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Symptoms</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Family_History</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Obesity</td>
<td>22</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>Previous_IFG_IGT</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Hypertension</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>HDL_Cholesterol</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Trigyceride</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Increased_Thirst</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Increased_Urine</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Increased_Appetite</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Weight_Variation</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Impaired_Vision</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Dataset as Knowledge Base

Machine Learning Model for Prediction of Diabetes Mellitus
Following table 2 contains the number representation for the symptoms, which are used for inehancing analysis.

### Table 2: Number Representation for Symptoms

<table>
<thead>
<tr>
<th>Boolean Value Symptoms</th>
<th>Multi Value Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family History</strong></td>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>True</td>
<td>Young 1</td>
</tr>
<tr>
<td>False</td>
<td>Adult 1</td>
</tr>
<tr>
<td></td>
<td>Old 1</td>
</tr>
<tr>
<td><strong>Obesity</strong></td>
<td><strong>Triglyceride</strong></td>
</tr>
<tr>
<td>Low</td>
<td>Normal 2</td>
</tr>
<tr>
<td>Normal</td>
<td>Boarder Line 3</td>
</tr>
<tr>
<td>Obese</td>
<td>High 4</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td><strong>HDL Cholesterol</strong></td>
</tr>
<tr>
<td>Normal</td>
<td>Very High 5</td>
</tr>
<tr>
<td>Elevatal</td>
<td>Low 1</td>
</tr>
<tr>
<td>High</td>
<td>Medium 2</td>
</tr>
<tr>
<td>Very High</td>
<td>High 3</td>
</tr>
</tbody>
</table>

IV. PROPOSED METHODOLOGY

The secondary data is collected from books, internet, medical journals and primary data collection done by providing questionnaires to physicians about symptoms affecting diabetes. The rules for this Machine Learning Model were designed after data analysis and expert suggestion. These rules are entered into CSV file which is dataset for the machine learning model has been used for further processing. Google Collabatory known as Colab is a cloud service based on Jupyter Notebooks is used for execution of machine learning algorithm [3][4] for this Machine Learning Model. This dataset contains the knowledge of the Machine Learning Model in the form of rows with numbers which are symptoms of diabetes disease. This respective Machine Learning Model is predicting the type of diabetes which is patients suffering from. Here dataset with values which are rules and knowledge base of the Machine Learning Model. Some of the rules for including three categories of diabetes diagnosis that can be interpreted are as follows.

Family_History, Age 11, Obesity 22, Previous_IFG_IGT 1, Hypertension 0, HDL_Colesterol 0, Triglyceride 1, Increased_Thirst 1, Increased_Urineate 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, Recurrresnt_fungal_infectn 0, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VagnalMycoticInfectn 1, Rapid_Heart_Beam 1, Recurring_Gum_Inf 0, Impotency 1, high blood Pressure 0, Sleep_Walking 1, Makec_usual_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_Sroke 0, Poly_Ovary_Syndrome 0, LowbloodSugar_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0, Last Column as type 1 diabetes.

Family_History, Age 12, Obesity 23, Previous_IFG_IGT 1, Hypertension 5, HDL_Colesterol 3, Triglyceride 5, Increased_Thirst 1, Increased_Urineate 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, Recurrresnt_fungal_infectn 1, Nightmares 1, Seizures 1, Sadness 1, Unconsciousness 1, Numbness 1, VagnalMycoticInfectn 1, Rapid_Heart_Beam 1, Recurring_Gum_Inf 0, Impotency 1, high blood Pressure 0, Sleep_Walking 1, Makec_usual_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 1, Difficult_Concentrating 1, Dehydration 1, LackOfCoordination 1, Hist_Heart_Sroke 0, Poly_Ovary_Syndrome 0, LowbloodSugar_NewbornBaby 0, WaistSize02cmM88cmF 0, WaistHipRatio.9M.85F 0, Last Column as type 2 diabetes.

The CSV file with first 64 columns with value of symptoms for rules and prediction value as last column in the dataset as type of diabetes is saved in knowledge base of the Machine Learning Model. After executing the machine learning algorithm on this Machine Learning based expert System will Predict the diabetes type of entered symptoms on command line.

V. MACHINE LEARNING BASED MODEL

Decision tree classifier is a systematic approach for multiclass classification. It poses a set of questions to the dataset. The decision tree classification algorithm can be visualized on a binary tree. On the root and each of the internal nodes, a question is posed and the data on that node is further split into separate records that have different characteristics.
The leaves of the tree refer to the classes in which the dataset is split. In the following code snippet, researcher train a decision tree classifier in scikit-learn. DecisionTreeClassifier is a class capable of performing multi-class classification on a dataset.

**Importing Libraries and Dataset:**
The following script imports required libraries; since our file is in CSV format, researcher used panda's read_csv method to read our CSV data file (figure 1)

Getting dataframe in the dataset and saving the labels into a variable. In the data analysis execute the df.head( ) and df.shape command for getting the number of rows and columns in our dataset. Prepare the data for further operation. Here divided the data into attributes and labels and will then divide the resultant data into both training and test sets. By doing this it can train the algorithm followed by testing. This provides a more accurate view of how your trained algorithm will actually perform. In the reprocessing step divide data into training and test sets. The model_selection library of Scikit-Learn contains train_test_split method will randomly split the data into training and testing sets. The data has need to be divided into the training and testing sets, the final step is to train the decision tree algorithm on this data and make predictions.

The predict method of the DecisionTreeClassifier class is used for making prediction. Over fitting can become drawback for making the knowledge base this problem gets solved by use of suitable algorithm and check for the alternative solution.

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

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

Figure 1: Importing Libraries and Dataset

VI. RESULT AND DISCUSSION

The proposed expert system has been validated on 10 persons. On the basis of symptom of patient the function dtree_model.predict( ) will take parameter. These parameters are numbers provided in table 1 and table 2. The symptoms based on which machine learning based expert system will execute prediction for the diabetic type. Five different prediction cases validated for 5 patients have been given below. After gathering of patient symptoms data this Machine learning Based Expert System produces the output for the patient in form of the diabetes type which the patient is suffering form.

During this research, a broad review of some models and expert system for diagnosing diabetes was carried out in other to identify basic comparisons between the proposed system and the already existing ones. Onuiri Ernest E et al proposed system was web based application used to diagnose diabetes by specifying symptoms and other parameters which aims to simulate an expert system. The application uses input parameters to determine the diabetes type[5]. Whereas Dilip Kumar Choubey et al designed a rule based fuzzy expert system which is clinical decision support system for diabetes for expediting diagnosis of diabetes, this rule based expert system is able to classify the type of diabetes into type 1, type 2, pre diabetes, gestational based on the presented symptoms. Through this proposed system will enhance the probability of diagnosing diabetes disease more accurately[6], comparatively our machine model using some more symptoms and the time, efforts for the designing the interfaces will reduced by use of machine learning platform. Tawfik Saeed Zeki et al used VP_Expert Primer for coding the expert system for diabetes. Their decision table of diagnosis which shows various combinations of patient’s situation, patient’s age, symptoms, effective factors and tests and by analyzing them provides the final decision of diagnosis[7]. In same way Anindito Yoga Pratama et al designed android base mobile application which is expert system. This mobile application determines a person’s diabetes risk. It is actually just a tool for early prevention for diabetes [8]. Hanslal Prajapati et al They proposed Fuzzy Inference System (FIS) which improve the accuracy and to achieve better efficiency. They designed and implemented a Fuzzy Expert System for Diagnosis of Diabetes. Pima Indian diabetic dataset was used as dataset. Fuzzy Logic Controller was designed in Matlab Fuzzy Logic Toolbox for Diabetes Diagnosis which consists of 8 Linguistic Inputs & 1 output as DM[9].

There are many predictive research in the area of social science [10], healthcare and many more domains. The complications related to diabetes like diabetic neuropathy[11], retinopathy etc also has broad scope of applying machine learning applications[12]. As Prediction has never end. Accurate and better research always in demand. For more accuracy current prediction research[13]. Therefore future studies for finding more accurate symptoms and by using power of newer technology enhance the accuracy of prediction rate like anything. For prediction or diagnosis of diabetes and its types for patient by using the Machine Learning Model researcher provided five different real cases out of ten patient. It has achieved exact results as doctors.

**Prediction Cases:**

**CASE 1:** Passing the values of symptoms for the first patient referred from Table 1 and Table 2. My_Pred=dtree_model.predict([[0,1,1,22,0,0,0,0,1,1,1,0,0,1, 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, 1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0]])

**CASE 2:**
Figure 2: Case 1 Prediction as Type1Diabetes
Output: Type1Diabetes

CASE 2:
Passing the values of symptoms for the second patient referred from Table 1 and Table 2.

Figure 3: Case 2 Prediction as Type2Diabetes
Case 2 Output: Type2Diabetes

CASE 3:
Passing the values of symptoms for the third patient referred from Table 1 and Table 2.

Figure 4: Case 3 Prediction as Type2Diabetes
Case 3 Output: Type2Diabetes

CASE 4:
Passing the values of symptoms for the fourth patient referred from Table 1 and Table 2.

Figure 5: Case 4 Prediction as Type2Diabetes
Case 4 Output: Type2Diabetes

CASE 5:
Passing the values of symptoms for the fifth patient referred from Table 1 and Table 2.

Figure 6: Case 5 Prediction as Type2Diabetes
Case 5 Output: Type2Diabetes

VII. CONCLUSION
This is an innovative technique from which researcher combines the power of the machine learning technique with the expert system. Here simple decision tree classification algorithm which is non-parametric supervised learning method used for
making the classification diabetes types. This model predicts the diabetes type of a target symptoms by learning simple decision rules inferred from the data features. This is the generalized method from which one can convert expert systems to machine learning platform. Regular and good research is always in demand, which help for more accurate prediction in area of Healthcare.

ACKNOWLEDGMENT

The authors thank to Kadsiddheshwar Swamiiji and team of physician at Siddhagiri Hospital & Research Center, Kaneri, Kolhapur for their support and Dr.R.K.Kamat, coordinator, Department of Computer science, Shivaji university for his assistance and motivation. Department of Computer science, Shivaji university for their infrastructure support.

REFERENCES


AUTHORS PROFILE

Ms. Manjiri Mahadev Mastoli pursued Master of Computer Application from CSIBER, Shivaji University, Kolhapur, Maharashtra. She is currently pursuing Ph.D. and currently working as research Scholar in Department of Computer Sciences, Shivaji University, Kolhapur, Maharashtra. Her main research work focuses on Artificial Intelligence and Machine Learning, Data Science. She has 5 years of teaching experience and 5 years of Research Experience.

Dr.Urmila Pol pursued Bachelor of Science from Shivaji University and Master of Computer Application from CSIBER, Shivaji University, Kolhapur, Maharashtra. She is currently working as Assistant Professor in Department of Computer Sciences, Shivaji University, Kolhapur, Maharashtra. She is Member of International Association of Engineer. Currently, she is guiding four Ph.D and one M.Phil student. Her main research work focuses on Artificial Intelligence, Data Science, Big Data Analytics, LMS and Open Source Technologies, She has 20 years of teaching experience and 15 years of Research Experience.

Mr. R.D. Patil pursued Bachelor of Engineering Shivaji University, Kolhapur and Master of Business Administration from IGNOU. He is currently currently working as Quality Assurance Engineer at Menon Bearing Ltd, Kolhapur, Maharashtra. His main research work focuses on Operational Research, Artificial Intelligence and Machine Learning. He has 8 years of Quality Assurance experience and 4 years of Research Experience.