e-ISSN: 2349-2678, p-ISSN:2961-6050

available online on www.ijpba.in

International Journal of Pharmaceutical and Biological Science Archive

NLM (National Library of Medicine ID: 101732687)

Index Copernicus Value 2022: 72.59

Volume 12 Issue 6; 2024; Page No. 159-169

To Study the Application and Role of Artificial Intelligence and Machine Learning in Healthcare

Mr. Hemant Patil[1], Dr. Deepak Kumar Birla[2]

[1]PhD. Scholar, Sunrise University, Alwar, Rajasthan

[2]Professor, Sunrise University, Alwar, Rajasthan

Article Info: Received: 17-09-2024 / Revised: 14-10-2024 / Accepted: 11-11-2024

Address for Correspondence: Hemant Patil

Conflict of interest statement: No conflict of interest

Abstract

The prevalence of diabetes has been steadily rising, and the illness is now recognized as a leading cause of mortality on a global scale. While diabetes cannot be cured, it may be treated and controlled with prompt identification. Combining AI with ML allows for automated early diabetes identification, which outperforms manual methods by a significant margin We present a novel Machine Learning based Model (MLM) to increase efficiency and enhance forecast accuracy. The patient's diabetes type, as well as whether it is type 1, type 2, or gestational diabetes, may be predicted by this Machine Learning Model (MLM). Innovative and more accurate than previous methods, the suggested Machine Learning Model may diagnose diabetes. As a significant cause of blindness on a global scale, diabetic retinopathy (DR) is a problem that is only becoming worse.

Keywords: Healthcare, Artificial Intelligence, diabetes and Machine Learning

Introduction

When it comes to patient care and the management of clinical data, machine learning is a lifesaver. It is a use of AI, which is the study and practice of teaching computers to do tasks normally performed by humans. has many potentials uses in healthcare, including but not limited to: managing patient records, seeing patterns in healthcare, making treatment recommendations, and more. New and interesting job possibilities have arisen as a result of hospitals and healthcare organisationsrealising that machine learning may enhance decision-making and decrease risk in the medical profession.

International Data Corporation (IDC) predicts that the value of India's artificial intelligence (AI) industry would more than double from 2020 to 2025. Startup activity, interest in AI, and job openings in the sector are all expected to rise as a result. The use of machine learning in healthcare is a rapidly developing and, surprisingly, accessible area. The concepts of "artificial intelligence" and "machine learning" may appear complex at first, but they really only need a basic understanding of mathematics and computer programming. You can tackle more complex ideas and problems using machine learning when you master its fundamentals. This has the potential to reveal untapped avenues for healthcare innovation and diversity in the workplace.

Large medical organisations have also begun to use machine learning-based approaches to better organise their electronic health records, detect abnormalities in organs, bones, and blood samples through medical imaging and monitoring, and to facilitate robot-assisted surgeries. The fight against COVID-19 has lately accelerated testing and hospital response thanks to machine learning techniques. In spite of the epidemic, hospitals have used the Clinical Command Centre, a deep learning system developed by GE, to manage, communicate, and monitor everything from patients and beds to rooms and ventilators to electronic health records and even employees. In addition to tracking SARS-CoV2 cases, researchers have used AI to identify genetic sequences, develop vaccines, and more.

LITERATURE REVIEW

Xu *et. al.* (2016) I have shown the use of a deep learning technique called Stacked Sparse Auto-Encoder (SSAE) for detecting nuclei in breast cancer histopathology pictures. SSAE is capable of addressing the challenges associated with variations in the size, shape, texture, and appearance of nuclei by learning the differentiating characteristics from the pixel intensities. The sliding window technique is conducted to each individual picture to extract high-level characteristics. These features are then inputted into the classifier to distinguish between nuclear and non-nuclear patches of the image. The given approach of nuclei identification, known as SSAE, outperforms previous state-of-the-art techniques, resulting in an improved F-measure of 84.49% and an average area under the precision-recall curve of 78.83%.

Aubreville*et. al.* (2017) We have introduced a novel approach for diagnosing oral cavity cancer by using deep learning algorithms on a dataset of 7894 Confocal Laser Endomicroscopy (CLE) pictures. The patch probability fusion approach is used to achieve effective picture categorization by focusing on specific patches rather than the full image. This approach is designed to meet the CNN's need for a bigger dataset and to reduce the number of learning parameters required to solve pattern recognition problems. In addition, the system successfully minimises the total classification error, resulting in a mean accuracy level of 86.6% and an area under the curve (AUC) of 0.96.

Lee *et. al.* (2016) In this study, two ways have been discovered. The first strategy involves applying a pooling function by merging the max and average pooling characteristics. The second technique involves the acquisition of knowledge about the pooling operation, which is then implemented via the utilisation of values inside the pooling filters. These learnt filters are integrated responsively to provide an extended kind of learning. These implemented solutions provide an improvement in performance and invariance characteristics compared to traditional pooling, with only a little increase in computational workload and a moderate increase in the number of parameters during training. These tactics are simple to implement and may be utilised in different deep network designs.

He *et. al.* (2017) I have presented a novel strategy for channel pruning to enhance the training pace of deep convolutional neural networks (CNNs). This method is meant to exploit and prune redundant channels using a two-stage process. In the first phase, the redundant channel included within feature maps is used and pruned using the Least Absolute Shrinkage and Selection Operator (LASSO) regression approach. In the second stage, the outputs are reconfigured from the remaining channels using linear least squares. This strategy's effectiveness is evaluated on three well recognised networks: VGG16, ResNet-50, and Xception-50. The training phase of the model is appropriately sped up by a factor of 4 for VGG16, resulting in a targeted improvement of 1% in the top-5 error rate. However, this approach speeds up the training process by a factor of 2 for ResNet-50 and Xception-50, but with a decrease in accuracy of 1.4% and 1% respectively.

Bardou *et. al.* (2018) We have used a Convolutional Neural Network (CNN) to achieve the multi-classification of breast cancer histopathology pictures using the BreaKHis dataset. The study conducted a comparative analysis between CNN and two handcrafted feature extraction approaches, Dense Scale Invariant Feature Transform and Speeded-Up Robust Features (SURF). These approaches were encoded using a bag of words and locality constrained linear coding, and further differentiated using SVM. This study illustrates the influence of using CNN as a feature extractor and its combined effect with handmade features. Furthermore, the analysis additionally considers the influence of the data augmentation approach on the magnification factor. It is found that there is a substantial decrease in the network's performance level specifically for a 40x magnification. The author, however, said that the accuracy level for binary classification ranged from 96.15% to 98.33%, while for multi-classification it ranged from 83.31% to 88.23%.

RESEARCH METHODOLOGIES

Records from references, the internet, and medical research publications make up the secondary data. Questionnaires about diabetic symptoms were sent to doctors as the main means of gathering data. Data analysis and expert recommendation went into designing the rules for this machine learning model. A CSV file, serving as the dataset for the machine learning model, has these rules written into it for further processing. The cloud infrastructure built on top of Jupyter Notebooks; Google Collaboration acknowledged by This Machine Learning based Expert System uses the Colab platform to execute machine learning algorithms. The machine learning model's information is stored in this dataset as a collection of numerical rows representing diabetic symptoms. The specific model or expert system in question is making diabetes type predictions for patients. Here is a dataset containing values that make up the expert system's knowledge base and rules.

DATA ANALYSIS

Machine learning model:

The most effective and methodical approach to multiclass classification is to use a decision tree classifier. The data set is queried using this positions list. Decision tree categorization may be best represented by a binary tree. At each level of the decision tree, questions are presented and the node is subsequently split to data with characteristics that are distinct. The depart node is referencing the class to which it belongs whenever dataset is divided or split. In order to train the decision_tree classifier from scikit_learn, researchers used the following codes. We employ the class Decision Tree Classifier, which has superior capacity and performance ability, for the multi_class classification of the data_set.

Importing Libraries' Datasets

Researchers use functional libraries to execute programme code correctly; the following code assists with importing libraries.see figure 1.

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

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

Fig 1: Importing Libraries and dataset

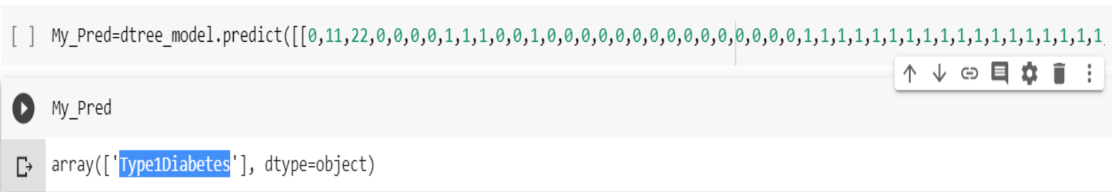
Retrieve the dataframe from the data set by storing the labels in a variable. To retrieve the count of rows and columns in the data set, run the data analysis using the df. head () and df. shape () commands. Get the data ready for what's to come. After sorting the data items into test and train sets according to their labels and attributes, the output is prepared. The technique may train the execution sequence and then test it. This is a detailed plan for how each algorithm will work. Data is divided in the reprocessing stage into a test set and a train set. The train-test-split() method is an integral part of Scikit_Learn's model selection library; it randomly divides the data_set into the train_test and test_set classes. Data must be divided into test and train modes before the algorithm can generate predictions in the final step. In order to generate forecasts, the Decision Tree Classifier class's predict function was used. The use of an appropriate method and the verification of alternative solutions eliminates overfitting, a potential negative of creating the knowledge base.

One hundred fifty people with diabetes served as validation subjects for this machine learning model. Out of 150 randomly selected test instances, this model has a 94% success rate. The dtree_model.Predict() method will accept a parameter based on the patient's symptoms. For the machine learning model to make appropriate judgements, it is based on the collection of signs and symptoms of diabetes persons, which are shown in tables 1 and 2. The following are five distinct prediction instances, each with verified results for five patients out of 150. This machine learning model takes the patient's symptom data and uses it to determine the kind of diabetes the patient has. The user then enters this information and gets the result.

Prediction Cases:

Case1:

Tables 1 and 2 are used to pass on the symptom values for the first patient. The following is the code for the dtree model: My Pred=predict ([[0,11,22,0,0,0,0,1,1,1,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0]]).



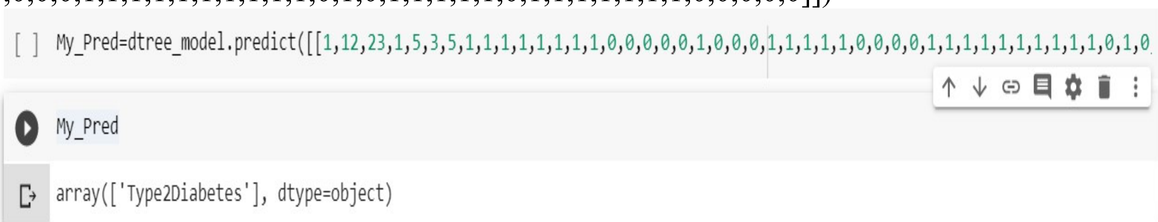
Case1: Diabetic Type 1 Prediction

Type 1 Diabetes Outcome

Case2:

PassingthevaluesofsymptomsforthesecondpatientreferredfromTable1andTable2.

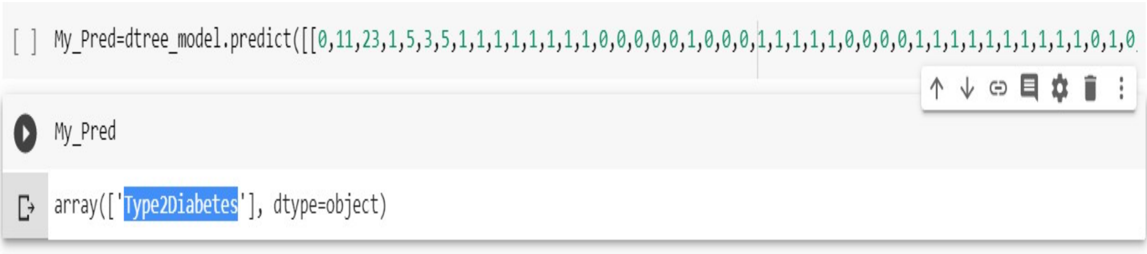
My_Pred=dtree_model. Predict ([[1,12,23,1,5,3,5,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,0,0,1,1,1,1,1 ,0,0,0,0,1,1,1,1,1,1,1,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,1,1,1,0,0,0,0,0]])



Case2The output of the prediction is type 2 diabetes.

Case3:

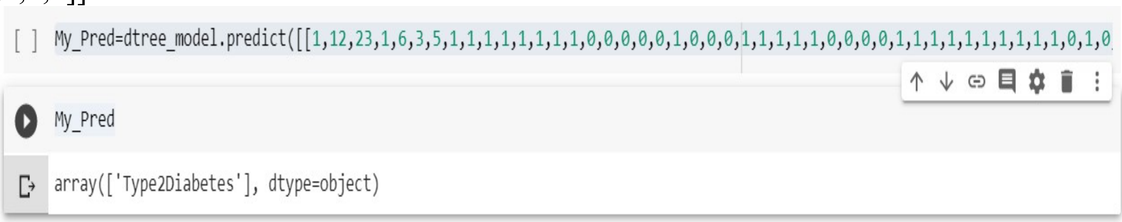
By transferring the symptom values from Tables 1 and 2 to the third patient. My Pred is the result of running the dtree_model and making a prediction based on the following data: [[0,11,23,1,5,3,5,1,1,1,1,1,1,1,1,1,1,1,0,0,1,0,0,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0 ,0,0]]



Case3Diabetic Outcome Prediction: Diabetic Outcome

Case4:

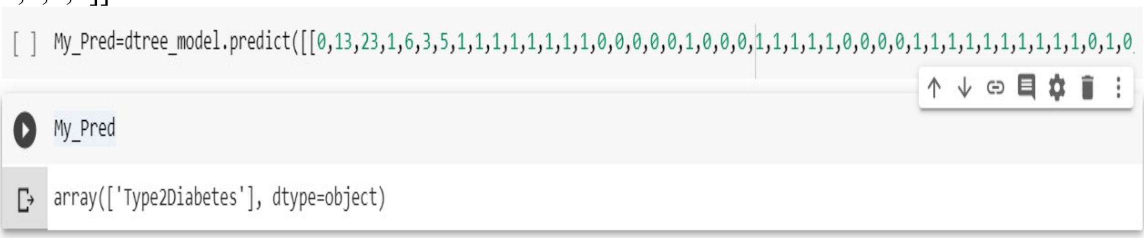
I am passing the symptom values from Tables 1 and 2 to the dtree_model. The prediction for the fourth patient is [[1,12,23,1,6,3,5,1,1,1,1,1,1,1,1,0,0,0,0,1,0,0,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,0,0,0,0,0]]



Case4Diabetic Outcome Prediction: Type 2 Diabetes

Case5:

I am passing the symptom values from Tables 1 and 2 to the dtree_model. The prediction for the fifth patient is [[0,13,23,1,6,3,5,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,0,0,0,0,0]]



Case5The output of the prediction is type 2 diabetes.

ENTROPY AND INFORMATION GAIN IN DECISION TREES

Entropy

According to Wikipedia, entropy is the state of chaos or uncertainty.

Entropy is defined as the measure of uncertainty or impurity in the set of instances.

- ***Entropy Working:***

Decision trees rely on a notion known as entropy, which has properties related to data separation, for their correct operation and construction.

A random variable's entropy may be defined as its degree of unpredictability, according to the researcher.

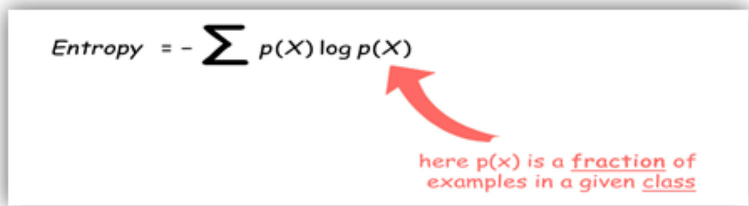


Fig 2: Calculation of entropy

Information gain with in Decision Tree:

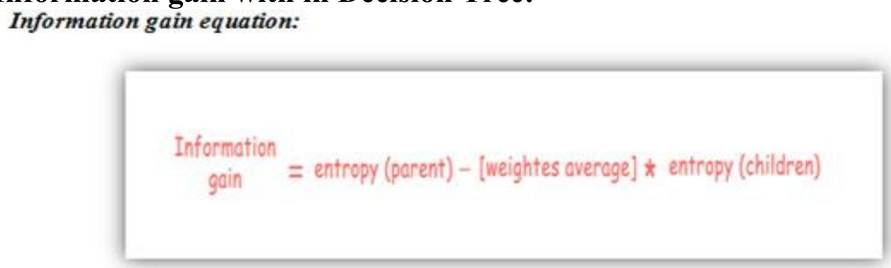


Fig 3: Calculation of information gain

Certainty Factor (CF) for Machine Learning Model:

Where there is ambiguity about the likelihood of a consequence, the primary approach to dealing with uncertainty is to employ certainty factors. One useful tool for ranking hypotheses according to relevance or importance is the certainty factor, which is a way to combine measures of belief and scepticism to a single number.

On the break, certainty considerations are clearly defined:

In this case,

$-1 \leq CF\ (H,\ E)\ \leq 1$

In this case, a value of zero represents no evidence, a number greater than zero supports the hypothesis, and a value lower than zero represents a negative hypothesis. The aforementioned CF values could be prompted by asking the following questions to patients:

How much do you believe that feeling' inceased Thirst' being diabetic?

If the evidence is to confirm the hypothesis

or

How much do you disbelieve that making a 'Slurred Speech' being diabetic?

For each question that has to be set, the value of the response is 80%.

CF (H | E) = 0.80 and

CF (H' | E) = -0.80

Experts often use words like "probably," "may," and "almost certainly" when assessing data. An uncertainty was placed using the certainty factor CF values.

Representation of the certainty factor:

CF [h,e]=M[h,e] - MD[h,e].............(I)

In this case, CF [h, e] represents the certainty factor.

The MB [h, e] value represents the confidence in the hypothesis 'h' given evidence 'e,' which might be anywhere from 0 to 1.

In MD [h, e], the values for evidence 'e' range from 0 to 1, and the magnitude of hypothesis 'h' is represented.

There are three possible outcomes for the hypothesis based on the available evidence. According to the theories, three things may happen based on the data that is now available. This is the first set of rules with a single piece of evidence (e) and a single hypothesis (h), and the representation of the certainty factor is:

CF[h,e]=CF(e) x CF(rule) ....(II)

While experts typically determined the certainty factor rule, patients were able to determine the CF (e) certainty factor by consulting the existing system. It is possible to perform a disjunction or conjunction based on the representation for the certainty factor that is computed using the interface that is used. Rules 2 with multi_evidence 'e' and a single hypothesis 'h' are shown below. The following is a description of the representation on connective disjunction in this case:

```
IF e1 AND e2 ... AND en THEN h (CF rule) ..........(III)
```

Where calculation of the value of CF combination, defined by:

```
CF(h,e)=min[CF(e1),CF(e2),...,CF(en)]*CF(rule)....... (IV)
```

In the meantime, for connecting conjunction, can be written as follows:

```
IF e1 OR e2 ... OR en THEN h (CF rule).............(V)
```

So the calculation of the value of the certainty factor (CF) combination, defined by the following representation:

```
CF[h,e] = max[CF(e1),CF(e2),...,CF(en)]*CF(rule).......(VI)
```

Last but not least, there's the one for the same hypothesis, which involves setting up two rules with different indications e1 and e2. The following are the details of the computations required to combine two certainty factors:

$$CF(CF1, CF2) = \begin{cases} CF_1 + CF_2(1 - CF_1) & if\ CF_1 > 0\ and\ CF_2 > 0 & \text{VII.A} \\ \dfrac{CF_1 + CF_2}{1 - [|CF_1|, |CF_2|]} & if\ CF_1 < 0\ or\ CF_2 < 0 & \text{VII.B} \\ CF_1 + CF_2(1 + CF_1) & if\ CF_1 < 0\ and\ CF_2 < (& \text{VII.C} \end{cases}$$

In this study, the researchers used method VII.A to calculate the rate of certainty factor twin evidence, as well as grouping and certainty factor calculations. However, for a single piece of evidence, they used formula four. The patients' certainty factor relations ranged from zero to one.

Patients were asked to write their certainty factor for each symptom during data collection, based on their views about that symptom. Some symptoms have a binary value that is either 0 or 1, and the certainty factor value ranges from -1 to 1, which is determined by the value of CF. (*values that do not apply)

A represents Type 2 Diabetes, B Type 1 Diabetes, and CFa and CFb, respectively, stand for the certainty factors for Type 2 and Type 1 Diabetes,

Table 3: dataset as knowledge base list of symptoms with CF values entered by patients

Sr. No.	Symptoms	A	CFa	B	CFb
1	Family_History	0	*	1	*
2	Age	11	*	12	*
3	Obesity	22	.90	23	.10
4	Previous_IFG_IGT	0	*	1	*
5	Hypertension	0	.90	5	.10
6	HDL_Cholesterol	0	*	3	*
7	Triglyceride	0	*	5	*
8	Inceased_Thirst	1	.95	1	.30
9	Increased_Urinate	1	.95	1	.30
10	Increased_Appetite	1	.90	1	.10
11	Weight_Variation	0	*	1	*
12	Impaired_Vision	0	.70	1	.50
13	Tiredness	1	.60	1	.40
14	Impatience	0	.50	1	.40
15	Infection	0	.30	1	.30
16	Itchy_Skin	0	.40	0	.40
17	Depression_Stress	0	.60	0	.60
18	Tingling_Sensation	0	.70	0	.70
19	Fruity_Breath_Odour	1	.40	0	.20
20	Bed_Wetting	1	.30	0	.30
21	Slow_Healing_Wound	0	.70	1	.70
22	FamilyHis_Pregnancy	0	*	0	*
23	Previous_Pregnancy	0	*	0	*
24	BabyOver_9Pd_PrePreg	0	*	0	*
25	Sleeplessness	1	.40	1	.30
26	Trembling	1	.40	1	.20
27	Sweating	1	.30	1	.20
28	Anxiety	1	.60	1	.50
29	Confusion	1	.20	1	.20
30	Weakness	1	.50	0	.30

31	Mood_Swings	1	.20	0	.10
32	Nausea	1	.10	0	.10
33	Vomiting	1	*	0	*
34	Dry_Skin	0	.10	1	*
35	Aches&Pains	0	.30	1	.30
36	Recurresnt_fungal_infectn	0	*	1	*
37	Nightmares	1	.40	1	.10
38	Seizures	1	*	1	*
39	Sadness	1	.10	1	.20
40	Unconsciousness	1	.10	1	*
41	Numbness	1	*	1	*
42	VaginalMycoticInfectn	1	*	1	*
43	Rapid_Heart_Beat	0	.20	1	.10
44	Recurring_Gum_Infe	0	.10	0	*
45	Impotency	1	*	1	*
46	high blood Pressure	0	.60	0	.60
47	Sleep_Walking	1	*	1	*
48	Makeg_unusual_noises	1	.10	1	.10
49	Leg_Cramps	1	.50	1	.20
50	Slurred_Speech	1	.10	1	.20
51	Flushed_face	1	.10	1	.10
52	Pale_Skin	1	*	0	*
53	LossofMenstruation	1	*	1	*
54	Stomach_Pain	1	.30	1	.10
55	Deep_Breathing	1	*	1	*
56	Areas_Darked_Skin	0	*	1	*
57	Difficult_Concentrating	1	*	1	.10
58	Dehydration	1	.30	1	.20
59	LackofCoordination	1	.10	1	.20
60	Hist_Heart_Stroke	0	.10	0	*
61	Poly_Ovary_Syndrome	0	*	0	*
62	LowbloodSugar_NewbornBab y	0	*	0	*
63	WaistSize02cmM88cmF	0	*	0	*
64	WaistHipRatio.9M.85F	0	*	0	*

Based on the user's or patient's chosen symptoms, the software aims to create an accurate profile by discovering or predicting the diabetes class that the user or patient is following.

Table 4: symptom with CF values entered by patients with predicted type of diabetes notify

CFno	Symptoms	A	CFa	B	CFb	A	B
CF8	Inceased_Thirst	1	.95	1	.30	✓	
CF9	Increased_Urinate	1	.95	1	.30	✓	
CF10	Increased_Appetite	1	.90	1	.10	✓	

The model predicts that the patient is suffering from type 2 diabetes based on data from table 4, where CF8 = 0.95 and CF9 = 0.95 are symptoms number 8 and 9, respectively.Therefore, a high CF value is required for type 2 diabetes.

Table 5: prediction using CF value

Predicted type of diabetes	CF value	Accuracy
Type2Diabetes	0.90	100%

The researcher has verified that the provided prognosis is accurate for all 50 cases. The certainty factors in an expert system that relies on machine learning are very valuable, but they are only appreciated after the system has performed its expert duties. The reliability of the model depends on this method. In the same way that an expert system can forecast the different forms of diabetes, this MLM can do the same.

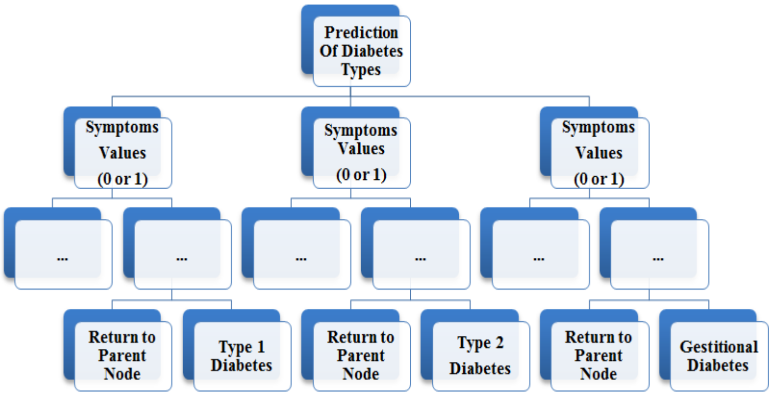


Fig 4: Decision tree for binary (0 or 1) values of symptoms

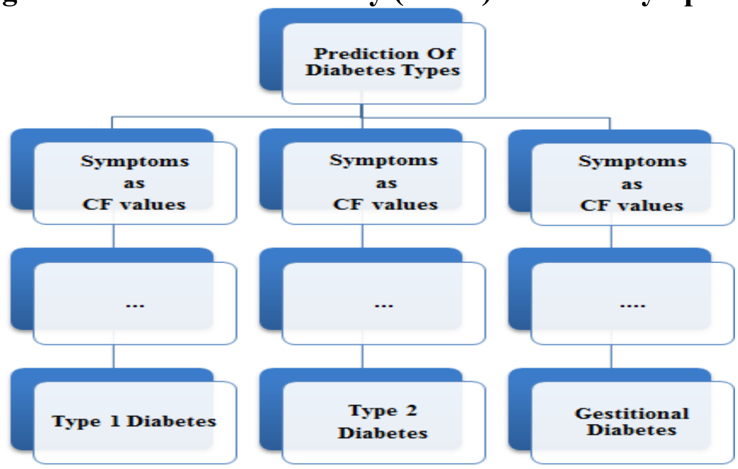


Fig 5: enhanced the classification with certainly factors

The patients are asked to record their level of confidence based on their views about their symptoms. In order to build the decision tree, the algorithm uses the confidence factor data, which is derived using the criteria of -1 and 1. This data represents the attribute for knowledge acquisition.

CONCLUSION

This article primarily aims to provide a concise summary of automated diabetes prediction using ML. The purpose of this study was to examine the several machine learning (ML) methods that have recently been created for the purpose of diabetes prediction and how well they work. Creating a diabetes prediction model aims to move the focus from increased accuracy to increased dependability for use in real-time settings. Few methods have trained and tested the model using many datasets. A global diabetes prediction model is required due to the rising prevalence of DM throughout the globe. We benefit much from machine learning and deep learning algorithms when they are used in healthcare. In comparison to current technologies, they are not only quicker and more efficient, but they also aid in improved illness prevention, diagnosis, and treatment.Based on what we know from previous research and our current findings, AI and ML techniques offer limitless potential uses in healthcare. Modern medical practice is benefiting from the use of AI and ML to streamline administrative tasks, provide more individualised care, and combat infectious illnesses.

REFERENCE

1. Xu, J, Xiang, L, Liu, Q, Gilmore, H, Wu, J, Tang, J &Madabhushi, A 2015, 'Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images', IEEE Transactions on Medical Imaging, vol. 35, no. 1, pp. 119-130.
2. Aubreville, M, Knipfer, C, Oetter, N, Jaremenko, C, Rodner, E, Denzler, J, Bohr, C, Neumann, H, Stelzle, F & Maier, A 2017, 'Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning', Scientific reports, vol. 7, no. 1, pp. 1-10.
3. Spanhol, FA, Oliveira, LS, Petitjean, C &Heutte, L 2016, 'Breast cancer histopathological image classification using convolutional neural networks', in 2016 international joint conference on neural networks (IJCNN), pp. 2560-2567.
4. Bayramoglu, N, Kannala, J & Heikkilä, J 2016, 'Deep learning for magnification independent breast cancer histopathology image classification', in 2016 23rd International conference on pattern recognition (ICPR), pp. 2440-2445.
5. Lee, C-Y, Gallagher, PW & Tu, Z 2016, 'Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree', in Artificial intelligence and statistics, pp. 464-472.
6. He, Y, Zhang, X & Sun, J 2017, 'Channel pruning for accelerating very deep neural networks', in Proceedings of the IEEE international conference on computer vision, pp. 1389-1397.
7. Bardou, D, Zhang, K & Ahmad, SM 2018, 'Classification of breast cancer based on histology images using convolutional neural networks', Ieee Access, vol. 6, pp. 24680-24693.
8. Akbar, S, Peikari, M, Salama, S, Nofech-Mozes, S & Martel, A 2017, 'Transitioning between convolutional and fully connected layers in neural networks', in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, pp. 143-150.
9. Howard, AG, Zhu, M, Chen, B, Kalenichenko, D, Wang, W, Weyand, T, Andreetto, M & Adam, H 2017, 'Mobilenets: Efficient convolutional neural networks for mobile vision applications', arXiv preprint arXiv:1704.04861
10. Anthimopoulos, M, Christodoulidis, S, Ebner, L, Christe, A &Mougiakakou, S 2016, 'Lung pattern classification for interstitial lung diseases using a deep convolutional neural network', IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207-1216.
11. Lakshmanaprabu, S, Mohanty, SN, Shankar, K, Arunkumar, N & Ramirez, G 2019, 'Optimal deep learning model for classification of lung cancer on CT images', Future generation computer systems, vol. 92, pp. 374-382.
12. Al-Tarawneh, MS 2012, 'Lung cancer detection using image processing techniques', Leonardo Electronic Journal of Practices and Technologies, vol. 11, no. 21, pp. 147-158.
13. Ada, RK 2013, 'Early detection and prediction of lung cancer survival using neural network classifier', International Journal of Application or Innovation in

Engineering and Management (IJAIEM), vol. 2, no. 6.
14. Arafi, A, Fajr, R &Bouroumi, A 2014, 'Breast cancer data analysis using support vector machines and particle swarm optimization', in 2014 Second World Conference on Complex Systems (WCCS), pp. 1-6.
15. Sanjay Dwivedi (2022): Study of investigative data mining techniques for combating terrorism http://hdl.handle.net/10603/368935