BioGecko

Vol 12 Issue 03 2023

ISSN NO: 2230-5807

Soft computing approach to build an Analytical and Predictive model thatanalyze the influence of gestational parents to juvenile diabetes

[1]Aparna Vaidyanathan Assistant Professor,
Department of Computer Science, Fergusson College, Pune.
[2] Dr. Vasantha Kalyani David Professor, Head of the Department,
Computer Science, Avinashilingam University,Coimbatore.

***Abstract*:** The diagnosis of type1 diabetes (T1D) in child has become a concern and a huge point of research with precision decisions. The main objective of the research is to make accurate model that apply complex algorithms to the datasets. As the lifestyle and environment changes the factors that influence diabetes is becoming more challenging and a scope of concern. A research is necessary to analyze the factors and the cause for type 1 diabetes in children that occurs in the early stage of age 4 to 7, in the age 10 to 14 and greater than 15. The concern is more critical where the diagnosis needs to be done much more early for the child that is detected with Type 1 diabetes as the child is dependent to insulin externally through injection. A Model that is analytical and predictive can assist in analyzing the occurrence of diabetes in children at the early stage so that prevention and cure can be provided at the right time with highest precision. The model emphasis on detecting the influence of gestational parents to the child in early stages. As there are various factors of influences, choosing the right factors and associating the gestational factors to the derivation of diabetes causes in the child can be made predictive and decisive with the extensive use of datasets. This research paper structures the predictive model to detect all possible influences that transmits diabetes to next generation from gestational parents using soft computing techniques. The objective of the paper focuses on accuracy and precision for the outcome of the analysis.

***Keywords*: Juvenile diabetes, soft computing techniques, gestational diabetes, Analytical model, influential factors, Predictive model, Algorithms.**

I. Introduction

Diabetes is a common chronic disease that is unpredictable and the reasons of occurrence becomes with various combination are a need of extensive research. Diabetes is characterized by hyper glycaemia and other metabolic abnormalities. In children and adolescents, one variant of the disease is common that is type 1 diabetes mellitus or juvenile diabetes. Juvenile is a **genetic disease** and children of parents with type 1 diabetes have a 30% chance of developing juvenile diabetes.During **pregnancy**, your placenta makes hormones that **cause** glucose to build up in your blood. Usually, the pancreas can send out enough insulin to handle it. But if the body can't make enough insulin or stops using insulin as it should, the blood sugar levels rise, and it detects **gestational diabetes**. The importance of criticality here is when a mother gives birth to a child in this situation, We require an analysis that predicts the possibility of a child with juvenile diabetes. The possibility includes various factors as it cannot be concluded with a fixed decision for every case. The model that is analytical and predictive will be the perfect tool to find the correlation between various factors that include in gestational parent and to identify the accuracy of transmission to the child with the age been grouped as less than 7, less than 14 and greater than 15.

In this paper we describe a methodology that assists the practitioners to prescribe the gestational T1D patients based on the factors that can be a cause to have a child with diabetes. An analytical and predictive model technique can provide a variant suggestion and analysis for the same. As the cause of juvenile diabetic has many factors that has numerous influences. The main factor is heredity and genetic parameter. Both are caused by a combination of **genetic** and environmental risk factors. However, there are other rare forms of **diabetes** that are directly **inherited**. The causes of **type 1 diabetes** are unknown, although several risk factors have been identified. Analysis of Juvenile diabetes data has still not been extensively researched in terms of gestational factors related to the causes. There is a lack of techniques applied to analyze the parent’s diabetes data and transformation of diabetes from gestational parents to child born. Hence there is a need of relative analysis on gestational data set to help in prevention of occurrence. The analysis for new-born babies who are Type 1 diabetic purely depend upon the gestational diabetic parent and with genetic factors.

The main objective of this research is to incorporate soft computing techniques to the data sets and analyze the correlation of gestational factors of the parent with the cause of diabetes to the child with the age group less than 15 years. The accuracy of the correlation between the gestational data and the juvenile data can be well arrived by applying analytical and predictive modelling techniques with soft computing approach for the given data set. The study also emphasis on the greater impact of gestational influences to the child. The system estimates the risk of diabetes in the patients by comparing its predicted results with patient's prior medical information. The paper emphasis on the impact and accuracy that Soft computing techniques can bring in the research outcome.

II. Literature Survey

This section reviews various research works that are related to the proposed work.

The research paper depicts Neural Networks are one of the soft computing techniques that can be used to make predictions on medical data. Neural Networks are known as the Universal predictors. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. The Artificial Neural Networks (ANNs) based system can effectively have applied for high blood pressure risk prediction. This improved model separates the dataset into either one of the two groups. The earlier detection using soft computing techniques help the physicians to reduce the probability of getting severe of the disease. The data set chosen for classification and experimental simulation is based on Pima Indian Diabetic Set from (UCI) Repository of Machine Learning databases. In this paper, a detailed survey is conducted on the application of different soft computing techniques for the prediction of diabetes. This survey is aimed to identify and propose an effective technique for earlier prediction of the disease. [1]

The research paper [2] As a chronic disease, diabetes mellitus has emerged as a worldwide epidemic. The aim of this study is to classify diabetes disease by developing an intelligence system using machine learning techniques. Our method is developed through clustering, noise removal and classification approaches. Accordingly, we use expectation maximization, principal component analysis and support vector machine for clustering, noise removal and classification tasks, respectively. We also develop the proposed method for incremental situation by applying the incremental principal component analysis and incremental support vector machine for incremental learning of data. Experimental results on Pima Indian Diabetes dataset show that proposed method remarkably improves the accuracy of prediction and reduces computation time in relation to the non-incremental approaches. The hybrid intelligent system can assist medical practitioners in the healthcare practice as a decision support system.

The highest incidence of type 1 diabetes worldwide, reaching 40 per 100 000 people per year in the 1990s. Our aim was to assess the temporal trend in type 1 diabetes incidence since 2000 in Finnish children aged younger than 15 years and to predict the number of cases of type 1 diabetes in the future. The incidence of type 1 diabetes in Finish children is increasing even faster than before. The number of new cases diagnosed at or before 14 years of age will double in the next 15 years and the age of onset will be younger (0–4 years). The Research paper mentions the highest rate of growth in Type1 diabetes in younger children. The findings also specify the age-specific rates per 100 000 per year were 31·0, 50·5, and 50·6 at ages 0–4 years, 5–9 years, and 10–14-years, respectively. We noted a significant non-linear component to the time trend ($p<0·0003$). In children aged 0–4 years, the increase was largest, at 4·7% more affected every year. The overall boy-to-girl ratio of incidence was 1·1; at the age of 13 years, it was 1·7 (1·4–2·0). The predicted cumulative number of new cases with type 1 diabetes before 15 years of age between 2006 and 2020 was about 10,800. [3]

Recent research hasbegun to emphasize children's adaptation and outcome in the larger context familyand other systems variables over time. The Riskand Resistancemodel, or the assertion that the effect of any risk factor that may vary depending upon other risk factors or buffering factors, has received empirical support. For instance, it is found that among some children with juvenile diabetes, the relationship between stressand blood glucose metabolic control was moderated by individual differences in copingwho received social support from both family and peersdemonstrated fewer behavior problems, as reported by their mothers. [4]

This research paper [5] emphasis on the Type 2 diabetes as frequently familiar in pregnancy might act in addition to genetic factors to cause diabetes in the children of mothers with gestational diabetes mellitus (GDM). The paper also identifies that Genetic predisposition to GDM should be equally shared by daughters of diabetic mothers and fathers. An excess of maternal transmission of diabetes is consistent with an epigenetic effect of hyperglycaemia in pregnancy acting in addition to genetic factors to produce diabetes in the next generation. The dataset taken in this paper shows the high influence of mother in the child to be affected with diabetes in early stage of life.

The purpose of this study was to determine the incidence of diabetes in women with previous dietary-treated gestational diabetes mellitus and to identify predictive factors for development of diabetes. Women with previous dietary-treated gestational diabetes mellitus have a considerably increased risk of later having diabetes. Follow-up investigations are therefore important, especially in those women with previous gestational diabetes mellitus in whom the identified predictive factors are present. [6]

III. Proposed Methodology

This paper focuses on how soft computing algorithms accelerate the analytical and Predictive model to contribute for predicting the children with diabetes with the factors related to gestational parents. The dataset that is used for analyzingproducesa inference rule of parameters that has a high impact on deciding the occurrence of diabetes in the patients. When there are large datasets and the rules for prediction are unclear mining techniques are used and the model is trained using the available datasets. The accuracy of the model is always a concern. The effectiveness of the algorithm is the measured through the accuracy of the algorithm with the data. In the case of predicting Type1 Diabetes, though physicians were able to intuitively estimate the risk of the disease, they could notanalyze the factors that can be relied upon. Machine learning algorithm will make the model to learn with new findings that can assist the physicians to analyze the factors which will assist them for diagnosing. While towards constructing the Machine Learning model, there are many approaches that can build the system,

This paper focuses on applying the soft computing algorithm to analyze the factors for Type1 Diabetes as majority of the data will have a pattern of categorical labels. There are many soft computing algorithms and tools that can be used.

Given a data set, this methodology can classify the data with the given set of values, the model can predict if the person will be prone to Type1 Diabetes symptoms and with what severity. Hence the prediction becomes a binary (yes/no) classification problem.

From a computational standpoint, classification problems are easier and more efficient to solve than any other data model. So, the research focuses on creating a classification model to transform the data to analyze the factors that will influence the medication of juvenilediabetic.

The gestational factors that can influence the juvenile's include the history of the parents, environmental factors, age, life- style ,overweight, obesity, gestational weight gain,HBA1c etc. When we apply Predictive alogorithms by deriving correlation between the factors and the target value of identifying the highest correlation values predicts whether the patient can have a diabetic child or not. The system learns with variant data sets and provide the decision based on the given dataset.

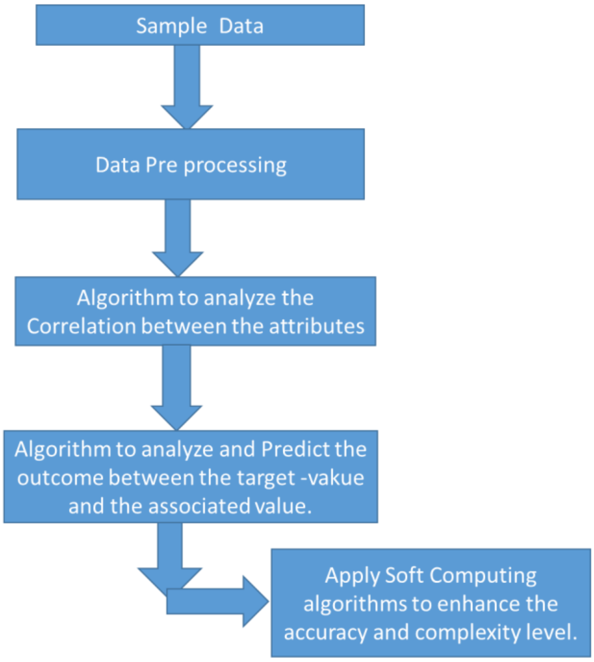


Figure1: Proposed Flow Diagram

Attribute Set		
1. Age	9. Other Disease	17. How taken
2. Sex	10. Adequate Nutrition	18. Family history affected in Type 1 diabetes
3. Area	11. Education	19. Family history affected in Type 2 diabetes
4. HBA1c	12. Standardised growth rate	20. Hypoglycemis
5. Height	13. Standardized birth weight	21. Pancreatic disease affected in child
6. Weight	14. Auto Antibodies	22. Affected (Target Variable)
7. BMI	15. Impaired glucose	
8. Duration	16. Insulin taken	

Table 1: Data set of Factor Table of Gestational Parent and Child.

Considering Table 1 the classification algorithm will generate rules of inferences that helps the practitioner to identify whether the gestational diabetic parents are likely to produce a new born with diabetic condition or prone to diabetic condition. In fortunate cases the new born can be treated and prevented from juvenile diabetic conditions by giving early medication to the parent.

3.1 Proposed Process of the Model

The data set of gestational factors with a set of 21 attributes is identified. An analytical and predictive model is built based on the factors identified in the data set. The model is trained to identify the highest proximity factor that can affect the child with diabetes. A direct proportionate and correlation is built between the factors and the target attribute. The factors are identified as Binary,Categorical , Nominal and Continuous variables. Algorithms are applied to analyse the relativity between the categorical attributes and the target variable. Here the target variable named “Affected” is binary with the values either Yes or No. The efficacy of the analysis is based on the factors that are selected with potential correlation of > .5 and the visualization of the prective analysis is based on the corelation value between the attribute and the target variable.

As per the decription of the data the algorithm of classification yields a accuracy of 95%. As the test data was increased the accuracy dropped to 75%. The computastional of the algorithm had to be increased with extreme complexity of Algorithm tuning and feature selection. Hence Soft Computing algorithms had to introduced to the data set.

The analytical and predictive model built on the given dataset projects the correlation of the attributes with the target variable. The model provides analysis and prediction based on the training set for the target variable as whether the child will be affected with diabetes Yes or No with the given attribute values.

3.2 Soft Computing Algorithms:

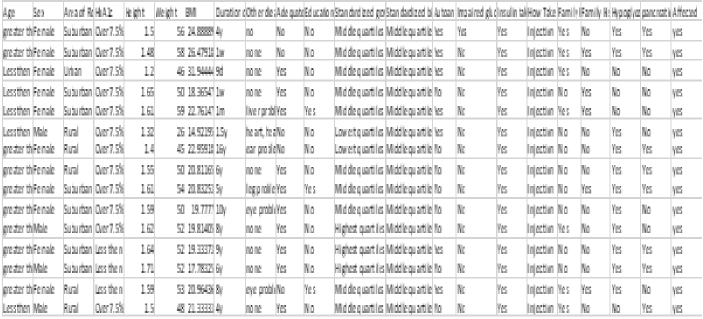
Artificial Neural Network The artificial neural network is much similar as natural neural network of a brain. Artificial Neural Network (ANN) basically has three layers, they are; Input layer: Input neurons define all the input attribute values for the data mining model, and their probabilities. Hidden layer: Hidden neurons receive inputs from input neurons and provide outputs to output neurons. The hidden layer is where the various probabilities of the inputs are assigned weights. A weight describes the relevance or importance of a particular input to the hidden neuron. The neuron with greater weight is assigned to an input. The value of that input is more important weights can be negative, which means that the input can inhibit, rather than favor, a specific result. Output layer: Output neurons represent predictable attribute values for the data mining model. [24]
Support Vector Machine: The standard model of supervised learning and test data are drawn from the same underlying distributions. This paper explores an application in which a second, source of data is available drawn from different distribution. The paper also introduces techniques for adjusting the kernel scores of the auxiliary data points to make them more comparable to the training data points [25]

Random Forest Algorithm:

Random forests are considered as a **highly accurate** and robust method because of the number of decision trees participating in the process. It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases.

3.3Sample data set:

As per **Table 1** Lets see the samples of the data set used to build the analytical and predictive model that predicts whether the child is affected to diabetes yes or no with the given gestational parent data.



Age	Sex	Area of Re	HbA1c	Height	Weight	BMI	Duration o	Other dise	Adequate	Education	Standardized gro	Standardized bi	Autoan	Impaired glu	Insulin tak	How Take	Family	Family Hi	Hypoglyc	pancreati	Affected
greater th	Female	Suburban	Over7.5%	1.5	56	24.88889	4y	no	No	No	Middle quartiles	Middle quartile	Yes	Yes	Yes	Injection	Yes	No	Yes	Yes	yes
greater th	Female	Suburban	Over7.5%	1.48	58	26.47918	1w	none	No	No	Middle quartiles	Middle quartile	Yes	No	Yes	Injection	Yes	Yes	Yes	Yes	yes
Less then	Female	Urban	Over7.5%	1.2	46	31.94444	9d	none	Yes	No	Middle quartiles	Middle quartile	Yes	No	Yes	Injection	Yes	No	No	No	yes
Less then	Female	Suburban	Over7.5%	1.65	50	18.36547	1w	none	Yes	No	Middle quartiles	Middle quartile	No	No	Yes	Injection	No	Yes	No	No	yes
Less then	Female	Suburban	Over7.5%	1.61	59	22.76147	1m	liver probl	Yes	Yes	Middle quartiles	Middle quartile	Yes	No	Yes	Injection	Yes	Yes	No	No	yes
Less then	Male	Rural	Over7.5%	1.32	26	14.92195	1.5y	heart, hea	No	No	Lowest quartiles	Middle quartile	Yes	No	Yes	Injection	No	No	Yes	No	yes
greater th	Female	Rural	Over7.5%	1.4	45	22.95918	16y	ear proble	No	No	Lowest quartiles	Middle quartile	No	No	Yes	Injection	No	No	Yes	Yes	yes
greater th	Female	Rural	Over7.5%	1.55	50	20.81165	6y	none	Yes	No	Middle quartiles	Middle quartile	No	No	Yes	Injection	No	No	Yes	Yes	yes
greater th	Female	Suburban	Over7.5%	1.61	54	20.83253	5y	leg proble	Yes	Yes	Middle quartiles	Middle quartile	No	No	Yes	Injection	No	Yes	Yes	Yes	yes
greater th	Female	Suburban	Over7.5%	1.59	50	19.7777	10y	eye probl	Yes	No	Middle quartiles	Middle quartile	No	No	Yes	Injection	No	No	Yes	No	yes
greater th	Male	Suburban	Over7.5%	1.62	52	19.81405	8y	none	Yes	No	Highest quartiles	Middle quartile	No	No	Yes	Injection	Yes	No	Yes	No	yes
greater th	Female	Suburban	Less then	1.64	52	19.33373	9y	none	Yes	No	Highest quartiles	Middle quartile	Yes	No	Yes	Injection	No	No	Yes	Yes	yes
greater th	Male	Suburban	Less then	1.71	52	17.78325	6y	none	Yes	No	Highest quartiles	Middle quartile	No	No	Yes	Injection	No	No	Yes	Yes	yes
greater th	Female	Rural	Less then	1.59	53	20.96436	8y	eye probl	No	Yes	Middle quartiles	Middle quartile	Yes	No	Yes	Injection	Yes	Yes	Yes	No	yes
Less then	Male	Rural	Over7.5%	1.5	48	21.33333	4y	none	Yes	No	Middle quartiles	Middle quartile	No	No	Yes	Injection	Yes	No	No	Yes	yes

https://www.kaggle.com/ dataset-of-diabetes-type1

Table 2: Sample Data Set.

The dataset in Table 3 has 15 sample records out of 306 records with 22 attributes.
The data was collected and made available from https://www.kaggle.com/ dataset-of-diabetes-type1. Several constraints were placed on the selection of these instances from a larger database. [7]

binary=['Adequate Nutrition','Education of Mother','Autoantibodies','Impaired glucose metabolism','Insulintaken','Family History affected in Type 1 Diabetes','Family History affected in Type 2 Diabetes','Hypoglycemis','pancreatic disease affected in child','Affected']
nominal=['Sex','Area of Residence','Other disease']
ordinal=['Age','HbA1c','Standardized growth-rate in infancy','Standardized birth weight','How Taken',]
count=['Height','Weight','BMI','Duration of disease']
attr=[binary,nominal,ordinal]

Table 3: Data type of Sample Data Set.

3.4. Brief Description of the Algorithm

The model is built using Python and some of its popular data science related packages. The data is read by Pandas to read our data from a CSV file and manipulate it for further use. The packages like NumPyis used to convert out data into a format suitable to feed the model. Visualization is done by seaborn and matplotlibmethods. Logistic Regression algorithm is imported from sklearn. This algorithm will help us build the model for prediction.

Step 1: Analysing the attribute with the target value and to find the highest correlated attribute with the target value.

Step 2: Apply Predictive algorithm to visualize the highest coorelated value and to make prediction of the target variable (i.e Affected with diabetes = (Yes/No)) for any given input variable of the dataset.

Here once the model is trained with the highest attribute coorelation and probability based on the training data set, the model is buit to predict the target attribute (i.e Affected with diabetes = (Yes/No))

Step 3: Apply soft computing technique for the given data set and check the accuracy measure improvement. Compare classification technque and softcomputing technique and measure the efficient accuracy.

3.5. Analysis of the Result:

As per the training dataset taken as input

Total record: 306.
Total Attributes: 22. (Input Attributes: 21, Target Attribute: 1)

I. Finding correlation of all attributes with 'affected'(target) attribute

1.	Affected (target variable)	1.000000
2.	How Taken	1.000000
3.	Insulin taken	1.000000
4.	HbA1c	0.754062
5.	Hypoglycemis	0.711744
6.	Duration of disease	0.660814
7.	pancreatic disease affected in child	0.578604
8.	Weight	0.575504
9.	Height	0.567638
10.	Age	0.530202
11.	Area of Residence	0.508678
12.	Autoantibodies	0.507387
13.	Adequate Nutrition	0.446675
14.	Other disease	0.434211
15.	Education of Mother	0.404526

16. Sex	0.287776
17. Standardized growth-rate in infancy	0.208690
18. Family History affected in Type 1 Diabetes	0.205045
19. Family History affected in Type 2 Diabetes	0.163438
20. Standardized birth weight	0.101381
21. BMI	0.038947
22. Impaired glucose metabolism	0.030192

Table 4: Correlation Attribute Table[7]

II. Analysing the data distribution.

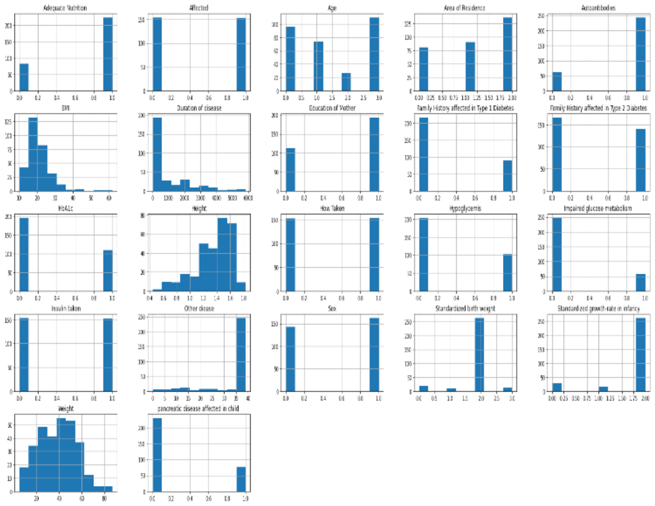


Figure 2 : Data Distribution[7]

III. Analysing the attribute with the target attribute where probability > .5 and directly proportional to the target value.

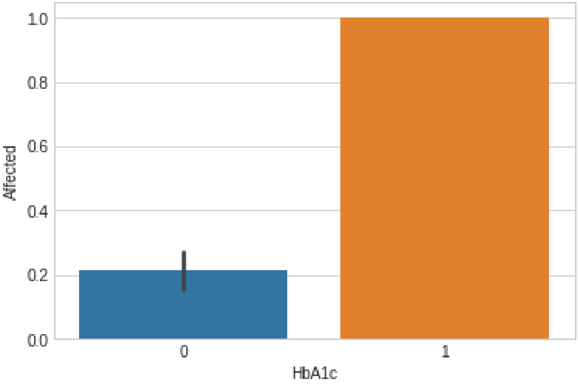
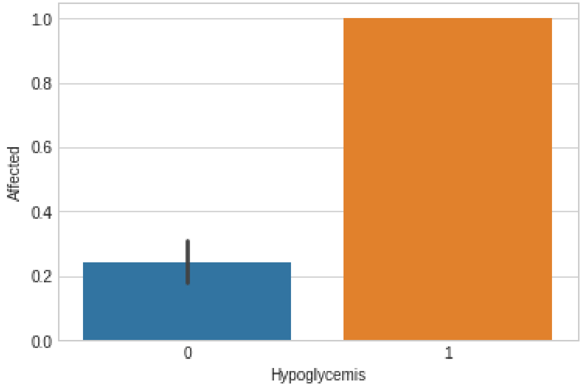


Figure 3: HbA1c attribute is direcly proportional with the target “Affected” attribute[7]



I.

Figure 4: Hypoglycemis is directly proportional with the target "Affected" attribute[7]

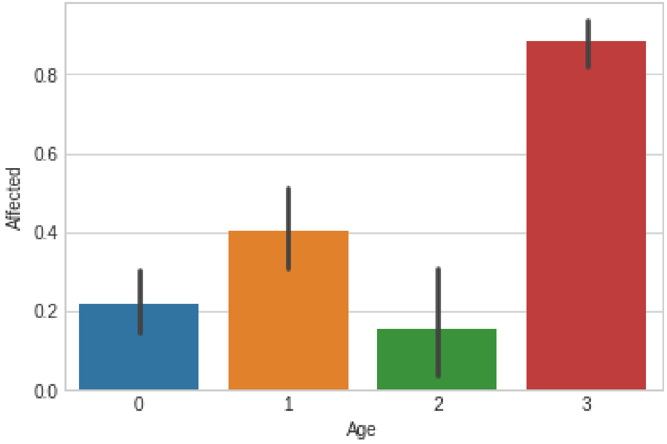


Figure 5: From above visualizations we can analyze that age with value 3(age greater than 15) is most likely to get affected by diabetes[7]

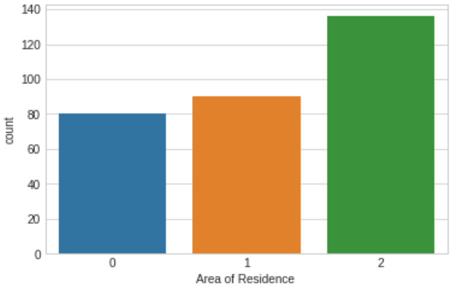


Figure 6: Area of Residence (This shows that People from rural area are most likely to get affected by Diabetes)[7]

As per the analysis Attributes **like Insulin Taken,HbA1c,Hypoglycemis, Age, Residence** have highest correlation with affected attribute.While attributes like **BMI and Impaired glucose metabolism** have lowest correlation.

3.6 Application of Soft Computing Algorithm:

3.6.1 SVM algorithm :

When the dataset is applied with SVM algorithm. The data was scrutinzed for non-dabetica and diabetc patients corelation with the given dataset. The **nondiabetic correlation** with gestational parents towards juvenile is plotted as:

Training Set= 60%
Test Set = 40%

Input Variable	Target Variable Diabietc ?	Corelation %
BMI	**N**	14%
Adeq Nutrition	**N**	1%

Table 5: Corelation Table for Non-diabetic data

Accuracy -= 95%

Random Forest Classifier

Random Forest is an Ensembled learning algorithm.

Support Vector Machine

SVM is an Machine alorithm that uses scatter plot for plotting the points.

Non Diabetic

Hb1Ac : 0
BMI : 14
Adeq Nutrition : 1
Hypoglycemia : 0
Type1 : 0
Type2 : 0
Growth Rate : 1
Birth Weight : 1
AutoAntibodies : 1
Impair Gluc : 0
Pancreatic Dis : 0

Accuracy

Predicted using SVM algorithm in sklearn.svm package.

Figure 7: SVM Algorithm Corelation Table for Non-diabetic data Accuracy -= 95%

The **diabetic correlation** with gestational parents towards juvenile is plotted as :

Training Set= 60%
Test Set = 40%

Input Variable	Target Variable Diabietc ?	Corelation %
BMI	**Y**	14%
Adeq Nutrition	**Y**	1%
Growth Rate	**Y**	2%
Birth Weight	**Y**	2%
Impair Gluc	**Y**	1%
Pancreatic Dis	**Y**	1%

Table 6: Corelation Table for diabetic data

Accuracy -= 95%

Random Forest Classifier

Random Forest is an Ensembled learning algorithm.

Support Vector Machine

SVM is an Machine alorithm that uses scatter plot for plotting the points.

Diabetes Predicted

Hb1Ac : 1
BMI : 14
Adeq Nutrition : 1
Hypoglycemia : 0
Type1 : 1
Type2 : 1
Growth Rate : 2
Birth Weight : 2
AutoAntibodies : 1
Impair Gluc : 1
Pancreatic Dis : 1

Accuracy

Predicted using SVM algorithm in sklearn.svm package.

Figure 8: SVM Algorithm Corelation Table for diabetic data Accuracy -= 95%

3.6.2 Random Forest algorithm:

When the dataset is applied with Random Forest algorithm. The data was scrutinzed for non-dabetica and diabetc patients corelation with the given dataset. The **nondiabetic correlation** with gestational parents towards juvenile did not reach to the desired outcome as the extension of data set was not more oriented to decision tree approach.

The **diabetic correlation** with gestational parents towards juvenile is plotted as :
Training Set= 60%

Test Set = 40%

Table 7: Corelation Table for diabetic data

Input Variable	Target Variable Diabietc ?	Corelation %
BMI	**Y**	14%
Adeq Nutrition	**Y**	1%
Growth Rate	**Y**	2%
Birth Weight	**Y**	2%
Impair Gluc	**Y**	1%
Pancreatic Dis	**Y**	1%
Type 1	**Y**	1%
Type 2	**Y**	1%
Antibodies	**Y**	1%

Accuracy -= 96%

Diabetes Predicted

Hb1Ac : 1
BMI : 14
Adeq Nutrition : 1
Hypoglycemia : 0
Type1 : 1
Type2 : 1
Growth Rate : 2
Birth Weight : 2
AutoAntibodies : 1
Impair Gluc : 1
Pancreatic Dis : 1

Accuracy

96%

Predicted using RandomForestClassifier algorithm in sklearn.ensemble.RandomForestClassifier package.

Figure 9: Random Forest Algorithm Corelation Table for diabetic data Accuracy -= 96%

IV. Result Analysis:

The model is further tested with test data and could learn that the model predicts exactly with accuracy of 95% with the training set = 60% and test set=40%. The test set is increased proportionally to make the model face more unpredictable input and the algorithm can be evaluated for its efficiency. The model is supported with an APP to take the input from the user and the model predicts whether the given parent based on his factors entered can have a child affected with diabetes Yes/No. The prediction is based on the learning of the model with the combination of 60% training test and 40% test set.

Algorithm	**Training and Test set %**	**Accuracy Measured**
Classification	70% - 30%	95%
SVM	60%-40%	95% (if Parent is Non-diabetic) 95% (if Parent Diabetic)
Random Forest	60%-40%	Minimal accuracy % (if Parent is Non-diabetic) 96% (if Parent Diabetic)

Table 8: Comparison of Algorithm with accuracy measure.

Conclusion

This study emphasis the gestational parameters that plays a vital role in transmitting juvenile diabetes to the next generation. The model developed using python programming based on the dataset emphasis on the influence of gestational parameters that include highest correlation has directly proportionate to children getting affected by juvenile diabetes. The accuracy of the model initially gets to 90% with training set and test set. The model was applied with 70% of training set and 30% test set to check with the accuracy of the model. The result gives the optimizedpredictive approach that proves the best of the model developed to Juvenile diabetic analysis. Further soft computing algorithm emphasis more on test set data so that the predictability can be verified. This emphasis more on machine learning algorithms that are complex and extensive to predict the transmission of diabetes from gestational parents to the children who are the next generation. The model helps the practitioner to prescribe the medication according to the suggestion from the application.

Further the research can include extensive datasets and complex machine learning algorithm to analyze and predict complex and unusual scenarios that occur in the dataset. This paper emphasizes that soft computing algorithms can handle the data set with more test data predictability than the classification algorithm. Further usage of Support Vector Machine algorithm can increase the accuracy than random forest algorithm by providing outcome for all varied parameters.

References

[1] International Journal of Scientific & Technology Research Volume 4, Issue 03, March 2015 ISSN 2277-8616 190 IJSTR©2015 www.ijstr.org Prediction Of Diabetes Using Soft Computing Techniques- A Survey M. Durairaj, G. Kalaiselvi

[2]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6177665/Childhood diabetes in india by Sanjay Karla and MuditaDhingare.

3] A soft computing approach for diabetes disease classification ***MehrbakhshNilashi*****,** ***Othman Bin Ibrahim*****,** ***Abbas Mardani*****,** ***Ali Ahani*****,** ***Ahmad Jusoh***
First Published November 14, 2016

[4]Predicting Diabetes Mellitus Using Artificial Neural Network Through a Simulation Study

[5] Predictive factors for the development of diabetes in women with previous gestational diabetes mellitus Author links open overlay panelPeterDammMD^{a}ClausKühlMD, PhDAkselBertelsenPhDLarsMølsted-PedersenMD, PhD

[6]Maternal Gestational Diabetes, Birth Weight, and Adolescent Obesity Matthew W. Gillman, Sheryl Rifas-Shiman, Catherine S. Berkey, Alison E. Field and Graham A. ColditzPediatrics March 2003, 111 (3) e221-e226; DOI: https://doi.org/10.1542/peds.111.3.e221

[7]A Predictive Model Approach to Detect the Influence in Transmission of Juvenile Diabetes from Gestational Parent to the Next Generation Assi. Prof. AparnaVaidyanathan1 , Dr. VasanthaKalyani David2 1Department of Computer Science, Fergusson College, Pune. 2Head of the Department Computer science, Computer Science, Avinashilingam University, Coimbatore.

[7]https://www.cs.usask.ca/faculty/kgs325/monitoring-human-behaviour/datasets/gestational-diabetes-datase.html
https://www.kaggle.com/ dataset-of-diabetes-type1

[8]Dr. Asir Antony Gnana Singh, Dr. Leavline, ShanawazBaig" Diabetes Prediction using Medicaldata", https://www.researchgate.net/publication/316432650_Diabetes_Prediction_Using_Medical_Data

[9]Sai PoojithaNimmagadda, SagarYeruva, Rakesh Siempu Improved Diabetes Prediction Model for Predicting Type-II Diabetes International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12, October, 2019

[10]Rabina1 ,Er. Anshu Chopra2 Research Scholar (CSE) 1, DIABETES PREDICTION BY SUPERVISED AND UNSUPERVISED LEARNING WITH FEATURE SELECTION

[11]SouadLarabi-Marie-Sainte 1,*, LinahAburahmah 1, RanaAlmohaini 1 and Tanzila Saba 2 Current Techniques for Diabetes Prediction: Review and Case Study Correspondence: slarabi@psu.edu.sa Received: 14 September 2019; Accepted: 21 October 2019; Published: 29 October 2019

[12]Tejas N. Joshi*, Prof. Pramila M. Chawan** Diabetes Prediction Using Machine Learning Techniques Int. Journal of Engineering Research and Application www.ijera.com ISSN: 2248-9622, Vol. 8, Issue 1, (Part -II) January 2018, pp.-09-13

[13]Perveen, S., Shahbaz, M., Guergachi, A., &Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. Procedia Computer Science, 82, 115-121. [13] Kamadi, V. V., Allam, A. R., &Thummala, S. M. (2016).

[14]A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach. Applied Soft Computing, 49, 137-145. [14] Pradeep, K. R., & Naveen, N. C. (2016, December).

[15]Predictive analysis of diabetes using J48 algorithm of classification techniques. In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I) (pp. 347-352). IEEE.

[16]Santhanam, T., &Padmavathi, M. S. (2015). Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. Procedia Computer Science, 47, 76-83.

[17] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., &Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, 29(2), 93-99.

[18] Aljumah, A. A., Ahamad, M. G., & Siddiqui, M. K. (2013). Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University-Computer and Information Sciences, 25(2), 127-136.

[19] Abdar, M., Zomorodi-Moghadam, M., Das, R., & Ting, I. H. (2017). Performance analysis of classification algorithms on early detection of liver disease. Expert Systems with Applications, 67, 239-251.

[20] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences, 29(2), 93-99.

[21] Bashir, S., Qamar, U., Khan, F. H., &Naseem, L. (2016). HMV: A medical decision support framework using multi-layer classifiers for disease prediction. Journal of Computational Science, 13, 10-25.

[22] Song, Y., Liang, J., Lu, J., & Zhao, X. (2017). An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing, 251, 26-34.

[23] Abdul-Ghani, M. A., &DeFronzo, R. A. (2009). Plasma glucose concentration and prediction of future risk of type 2 diabetes. Diabetes Care, 32(suppl 2), S194-S198.

[24] Prediction Of Diabetes Using Soft Computing Techniques- A Survey M. Durairaj, G. Kalaiselvi

[25] Improving SVM accuracy by training on auxiliary data sources. **Authors:** Pengcheng WuOregon State University, Corvallis, Thomas G. Dietterich

,