Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 18 Issue 4 Version 1.0 Year 2018
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Computation Model for Identifying Types of Diabetics using Multi-Selection Criteria Evaluation and K-Nearest Neighbor

By Towhidul Islam, Professor. Dr. Md. Nurul Mustafa & Jamir Ahmed
*Southern University Bangladesh*

*Abstract-* The main purpose of this topic is to develop a dynamic model of a Diabetes solution system. Diabetes mellitus is a chronic disease caused by inherited and/or acquired deficiency in production of insulin by the pancreas or by the ineffectiveness of the insulin produced such a deficiency results in increased concentrations of glucose in the blood. Which in turn damage many of the body systems in particular the blood vessels and nerves. Diabetes mellitus, often simply referred to as diabetes, is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced. This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphonic (increased hunger).diabetes is data mining based notification systems. This system developed is main purposed the people easily treatment for accommodation.

*Keywords: classification diabetes, MCES algorithm-NN algorithm.*

*GJCST-C Classification: I.5.m*

COMPUTATIONMODELFORIDENTIFYINGTYPESOFDIABETICSUSINGMULTISELECTIONCRITERIAEVALUATIONANDKNEARESTNEIGHBOR

*Strictly as per the compliance and regulations of:*

# Computation Model for Identifying Types of Diabetics using Multi-Selection Criteria Evaluation and K-Nearest Neighbor

Towhidul Islam [α], Professor. Dr. Md. Nurul Mustafa [σ] & Jamir Ahmed [ρ]

*Abstract-* The main purpose of this topic is to develop a dynamic model of a Diabetes solution system. Diabetes mellitus is a chronic disease caused by inherited and/or acquired deficiency in production of insulin by the pancreas or by the ineffectiveness of the insulin produced such a deficiency results in increased concentrations of glucose in the blood. Which in turn damage many of the body systems in particular the blood vessels and nerves. Diabetes mellitus, often simply referred to as diabetes, is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced. This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphonic (increased hunger).diabetes is data mining based notification systems. This system developed is main purposed the people easily treatment for accommodation. This system helps finds diabetes what type of diabetes type1, type2, & no diabetes easily provide this system. In this work, at first identity all the dependent variable or data to classify the suitable from unsuitable location. Then I have classified the data using Multi Criteria Evaluation System (MCES)[2].MCES helped the data set to be properly design and manipulated the system and K-Nearest Neighbor helped the diabetes range[1]. The main purpose of applying this identification is diabetes level. The concept of basically helped to build knowledge base. Most important of this topics collecting the real data for diabetes information.

*Keywords: classification diabetes, MCES algorithm-NN algorithm.*

## I. Introduction

Diabetes solution is data mining based notification systems. This system developed is main purposed every people easily known diabetes patient. Then what kinds of diabetes type 1,type 2 & No diabetes lives and easily find out for accommodation. This system helps easily finds out type 1, type 2 & Nodiabetes identify. This system designed basically Multi Criteria Evaluation Systems (MCES) method used[2]. MCES computing the data set to be properly design and manipulated the system. The main purpose of applying this identification is to design a physical level. This system works, at first select different types Symptoms. Example: frequently urination, very thirsty. weight less increased hunger tried and weakness injured dry delay every see present itch bred few see eye blurry vision irritability tingling gum infections etc. Second step in test selection example OGT, FGT, Hbalc, ABF, HDL, Blood etc[4]. Each Criteria select based on need then test result provide. Then base test value so result provide type1, type 2 or No diabetes. Next scaling the each criterion diabetes level using K-Nearest Neighbor, K-Nearest Neighbor helped the diabetes range. Then type 1 range sum>=130.55mm & Hbalc>=6.5 then type 2 diabetes range sum>=130.55mm & Hbalc<6.5 then type 2 diabetes and range sum<=130.55mm then no diabetes [6]. The main purpose of applying this identification is diabetes level. The concept of basically helped to build knowledge base. Most important of this topics collecting the real data for diabetes information. Standardization of criterion scores particularly as sign the value. All the value defined between two intervals scores o and 1.The maximum value is score 0, the minimum value is score 1, the mid value is score 0.5 and other value in scores 0 and 1[2].When the patient search symptoms then gives the measurable for each criteria particularly importance of this criteria .Finally, this system select a perfect test result provide then drug suggestion and dose time.

## II. Diabetes Mellitus

Diabetes mellitus is a chronic disease caused by inherited and/or acquired deficiency in production of insulin by the pancreas or by the ineffectiveness of the insulin produced such a deficiency results in increased concentrations of glucose in the blood. Which in turn damage many of the body systems in particular the blood vessels and nerves. Diabetes mellitus, often simply referred to as diabetes, is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced[4].

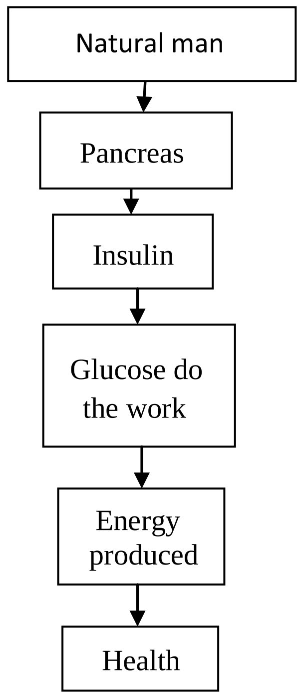
### a) *Because blood glucose very high*

This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia

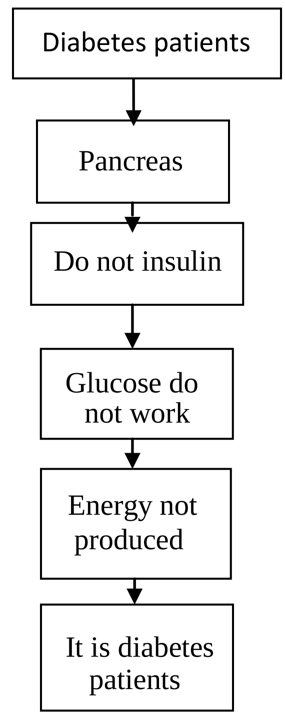
*Author α: M.Sc. in Computer Science & Engineering Southern University Bangladesh. e-mail: towhid.cse4@gmail.com*
*Author σ: Vice-Chancellor, Southern University Bangladesh. e-mail: mustafamn_bd@yahoo.com*
*Author ρ: Assistant Professor, CSE Department Southern University Bangladesh. e-mail: jahmedjbd@gmail.com*

(increased thirst) and polyphonic (increased hunger: under the design of process natural man and diabetes man[4].

Design of Process: Non -Diabetes




Design of Process: diabetes patients




b) *Symptoms*

Diabetes mellitus is a chronic disease caused by inherited and/or acquired deficiency in production of insulin by the pancreas or by the ineffectiveness of the insulin produced such a deficiency results in increased concentrations of glucose in the blood. Under the diabetes symptoms[4].

- Frequently urination
- Very thirsty.
- Weight less
- Increased hunger
- Tried and weakness
- Injured dry delay
- Every see present itch bred
- Few see eye
- Blurry vision
- Irritability
- Tingling
- Gum infections

c) *Reason for diabetes*

Which in turn damage many of the body systems in particular the blood vessels and nerves. Under the Diabetes reason for diabetes[6].

- Family father mother near relative to diabetes ineffective
- Weight very high
- Do not physical exercise and hardworking
- Longtime Cotswold medicine use

d) *Conditions that cause diabetes*

Diabetes mellitus is a chronic disease caused by inherited and/or acquired deficiency in production of insulin by the pancreas or by the ineffectiveness of the following conditions[6].

- Physical thickness
- Pregnant
- Injured
- Shock
- Surgery
- Mental contrariety

e) *Principle forms of diabetes*

i. *Type 1 diabetes:* insulin dependent in which the pancreas fails to produce the insulin which is essential for survival this form develops most frequently in children and adolescents but is being increasingly noted later in life. Type 1 diabetes, formerly called juvenile diabetes or insulin-dependent diabetes, is usually first diagnosed in children, teenagers, or young adults. In this form of diabetes, the beta cells of the pancreas no longer make insulin because the body's immune system has attacked and destroyed them. Treatment for type 1 diabetes includes taking insulin shots or using an insulin pump, making wise food choices, exercising regularly, taking aspirin daily (for some), and controlling blood pressure and cholesterol[6].

ii. *Type 2 diabetes:* non insulin dependent which results from the body inability to respond properly to the action of insulin produced by the pancreas. Type 2 diabetes is much more common and accounts for around 90% of all diabetes cases worldwide. It occurs most frequently in adults but is being noted increasingly in adolescents as well. Type 2 diabetes, formerly called adult-onset or non insulin-dependent diabetes, is the most common form of diabetes. People can develop type 2 diabetes at any age, even during childhood. This form of diabetes usually begins with insulin resistance, a condition in which fat, muscle, and liver cells do not use insulin properly. At first, the pancreas keeps up with the added demand by producing more insulin. In time, however, it loses the ability to secrete enough insulin in response to meals. Being overweight and inactive increases the chances of developing type 2 diabetes. Treatment includes taking diabetes medicines, making wise food choices, exercising regularly, taking aspirin daily (for some), and controlling blood presser. There are several signs and symptoms that indicate a person may have either pre-diabetes or undiagnosed diabetes[6].

## III. MCES (Multi Criteria Evaluation System)

MCES is basically called decision making process. MCES provides a framework for exploring solution to decision making problem, which may be poorly defined. It is a method for combining data according to their importance in making a given decision. At a conceptual level, MCES method involve qualitative or quantitative weighting, scoring or ranking of criteria to reflect their importance to either a single or a multiple set of objectives .The main advantage of MCES is that they make it possible to consider a large number of data, relations an objectives which are generally present in a specific real world policy problem, so that the problem at hand can be studied in a multi dimensional fashion. Perhaps the simplest MCES is the weighted linear summation system .The steps involved in applying this system a diabetes solution system is illustrated in Figure 01and can be described as follows.

*Step 1*

At first select Symptoms and different test of selection criteria.

*Step 2*

Standardization criterion scores of their measurable. Most MCES analysis, especially those using quantitative and mixed data sources, require some form of standardization of the scales of measurement used by the data layers. This is necessary to facilitate the comparison of factors measures using different units and scales of measurement.

*Step 3*

Allocation weighted of their each criterion. This is done by adding weighted to reflect the importance of each criteria. A high level of importance done maximum of weighted and low level of importance done minimum of weighted.

*Step 4*

Finally, applying the MCES method. An MCES method may then multiply theses standardized scores by the weights for each of the data layers in stage 1 and sum these to allocate a score to each pixel on the output map. Further evaluation of the results may be carried out by ranking the values in the results map and reclassifying the ranked map to show the top ranked correct symptoms or Test. This test easily indicates then kinds of diabetes easily provide.

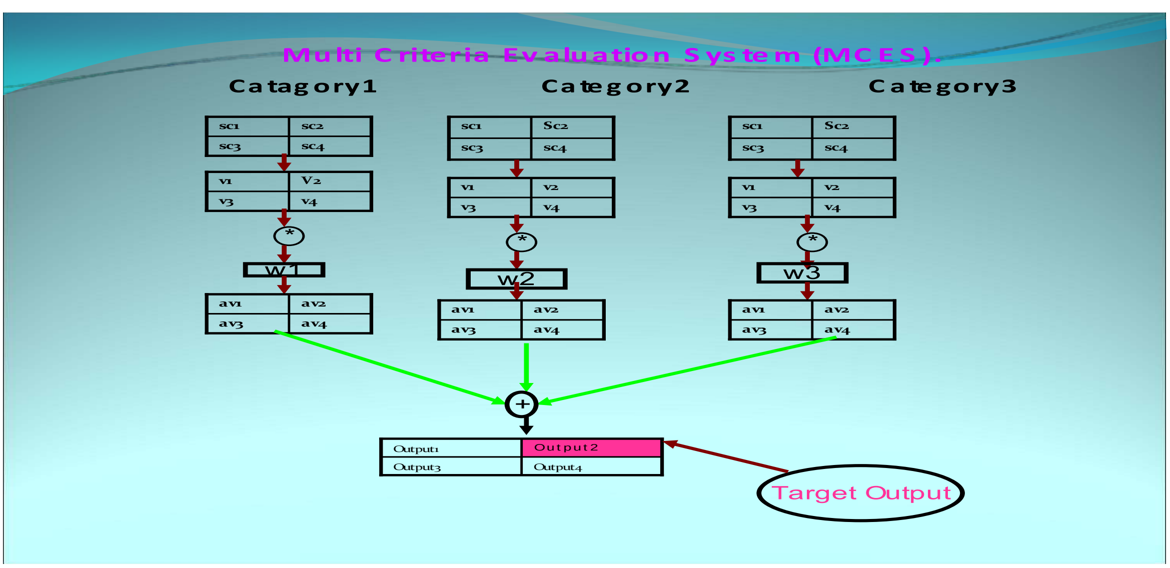



*Fig. 1:* Applying a linear weighted summation model in diabetes solution system

*a) K-Nearest Neighbor (K-NN)*

K–nearest neighbor (K-nn) algorithm is a branch of supervised learning[1]. Now-a-days it is being applying in various fields of data and information processing irrespective of science, commerce and arts. In the context of machine learning, K-nn is considered an effective data classification technique based on adjacent developed examples of sample space. The value of K is always positive and an object is classified by considering the greater number of choice of its neighbors. The neighbors are chosen from data set which is best fit for correct classifications and Euclidean distance helps to measure the overall distances. Here every occurrence correlates to points in sample space or within populations. Generally distance or similarity between instances or objects is easy if the data sets are numeric or integer. A very typical formula to calculate distances is Euclidian distances formula as follows:

$$d = \sqrt{} (x_{w1}-x_{w2})^2 + (x_{w21}-x_{w22})^2 + \ldots \ldots a$$

In some cases Manhattan or City Block distance also applicable:

$$d = (xw1-xw2) + (xwi-xwi2) + \ldots \ldots b$$

However it is very essential to bear in mind that all the instances at sample space must be same scale. As for example income will compare with income not the height of the human beings.

For qualitative data the distance measurement process will be different and it is important to consider that the instances are same or not. At this stage the qualitative objects are measured by allocating Boolean values to each object. It might be possible to converts to instances between which distance can be identified by some techniques. As for example color, temperature, age, height etc. Text and character has identified as one instance per word with the frequency start from 0, 1, 2……………………………n.

*b) The classifications process of K-nn as follows*

The two main steps of K-nn must follow are:

1. Training
2. Predictions

Training means to get information from all sample spaces and populations. To accomplish this work we need to have the idea about the all instances and objects. In this sense it is very important to bear in mind that data set must be in same class. The qualitative and quantitative data measurement will be different. The predictions will manage by considering the predefined methods.

i. *The k-nn Algorithm*

The total algorithmic steps are as follows:

1. Parameter selections (int m, int n). m=0, n=1, 2, 3………………………n.
2. Distance calculation

$$\sqrt{\sum_{i=0}^{n} (q-p_i)^2} \qquad \text{where } i=0,1,2,3\ldots\ldots\ldots, n$$

3. Short the distances of sample space and marked the closet neighbors in the context of K-th smallest distance.

Input instances *h* with *n sample objects*, comparator *sum*

Output instances *sum* sorted according to *h*

if *sum. length*() > 1

Then $(S_1, S_2) \leftarrow divide(S, n/2)$

SHORT NEIGHBORS $(S_1, C)$

SHORT NEIGHBORS $(S_2, C)$

$S \leftarrow$ SHORT NEIGHBORS $(S_1, S_2)$

4. *Similarities assumption:* Instances that are close together should have similar values.

Minimize

$$\xi(f) = \sum w_{ij}(s1-f_{s2})^2$$

Where $w_{ij}$ is the similarity between examples i and j. And $f_i$ and $f_j$ are the predictions for example i and j.

5. Predict the value as follows:

Standard KNN
$$\hat{y} = \arg\max_y C(y, Neighbors(x))$$
$$C(y, D') \equiv |\{(x', y') \in D' : y' = y\}|$$

6. Find out the best heuristics distance

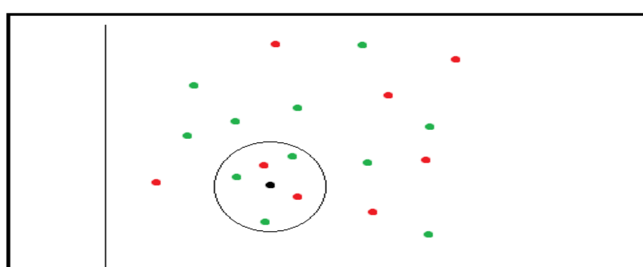
$$f(n) = g(n) + h(n)$$

Where:

- g(n) is the cost of the best path found so far to n
- h(n) is an admissible heuristic
- f(n) is the estimated cost of cheapest solution through n

*c) New Maximum Nearest Area (NMNA)*

How k-nn selects the desired values from a lot of alternatives is that it calculates its nearest most predicted value. The following figure depicts the computations.



5-NN-find 5 closest to black point, 3 Green and 2 red, So Predict Green

*Fig. 3:* The neighbor selecting process.

In the figure above we see that the small circle belongs three different color dots where the black one is the pivotal element and based on that point we will calculate the green and other two green and red points. According to this figure we have to predict the green points as a K nearest neighbors. The neighbors are very closest to the pivotal point.

*d) Organization of The Process*

Now it is important to build the process how K-nn may organized in reality or the time line. To manage the proper training area we have to shorten the area or to select the appropriate area. When we are able to fix the sample area for computation, it will help us to reduce the computational complexity for entire process.

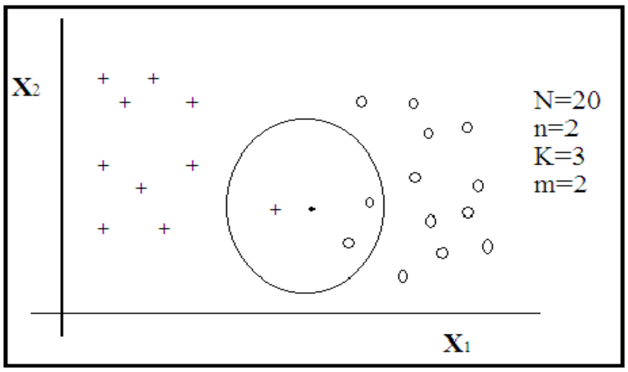



*Fig. 4:* The total population area for K nearest neighbors.

Here,

N= Total number of data set at population space. In this figure above we see that there is twenty (20) objects are outside the circle. The circle denotes the selected sample space. Inside the circle the black point indicate the pivotal or central point.

K=The total neighbors. Here the value of K is three (3).

n= indicate the nearest value.

m= categories of the neighbors. In the figure above we see that there are two categories of neighbors. One data set indicate by plus (+) sign and other is small hole.

## IV. Data Analysis

A Data analysis is integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of data referenced information. Data analysis allows us to view, understand, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of measurement, globes, reports, and charts. Its can be integrated into any enterprise information system framework. The integration of data which may have been obtained from various sources, computerized at various scales, and based upon different projection systems, is a complex task and remains a major challenge. In a general sense, the term describes any information system that integrates stores, analyzes, shares, and displays data information for informing decision making. Finally, its can produced different types information is combined relation each other.

Diabetes mellitus is characterized by recurrent or persistent hyperglycemia, and is diagnosed by demonstrating any one of the following as shown in the table (1) which is diabetes diagnostic criteria. Fasting plasma glucose level≥ 7.0 mmol/l (126 mg/dl) Plasma glucose ≥ 11.1 mmol/l (200 mg/dl) two hours after a 75 g oral glucose load as in a glucose tolerance test Symptoms of hyperglycemia and casual plasma glucose ≥ 11.1 mmol/l (200 mg/dl) Glycated hemoglobin (Hb A1C) ≥ 6.5%[7].

*Table 1:* diabetes diagnostic criteria

| Condition | 2 hour glucose | Fasting glucose | HbA 1c |
|---|---|---|---|
| Unit | mmol/l(mg/dl) | mmol/l(mg/dl) | % |
| Normal | <7.8 (<140) | <6.1 (<110) | <6.0 |
| Impaired fasting glycaemia | <7.8 (<140) | ≥ 6.1(≥110) & | 6.0–6.4 |
| | | <7.0(<126) | |
| Impaired glucose tolerance | ≥7.8 (≥140) | <7.0 (<126) | 6.0–6.4 |
| Diabetes mellitus | ≥11.1 (≥200) | ≥7.0 (≥126) | ≥6.5 |

*Step 1*

At first select symptoms and test of selection criteria. Most symptoms & test is selected based then kinds of diabetes provide. Example then Showing bellows:

*Table 1:* At first select symptoms and test of selection criteria.

| Kind of diabetes | Criteria 1 | Criteria 2 | ………. | Criteria 11 | Criteria 12 |
|---|---|---|---|---|---|
| Type1/Type2/No Diabetes | Thirsty | Weakness | ……….. | OGT | Hbalc |
| " | " | " | ………. | " | " |
| " | " | " | …….. | " | " |

*Step 2*

Secondly, scaling the each criteria particularly and assign the value or (measurable). Showing bellow:

*Table 2:* Scaling the each criteria particularly and assign the value

| Kind of diabetes | Thirst | Weakness | ……… | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 5 | 3 | ………. | 130 | 6 |
| Type2 | 1 | 0 | …….. | 150 | 10.5 |
| No diabetes | 2 | 1 | …….. | 135 | 00 |

*Step 3*

Standardization of criterion scores particularly distances value. All the value defined between two intervals scores o and 1.The maximum value is score 0, the minimum value is score 1, the mid value is score 0.5 and other value in scores 0 and 1. This score create the based on equation (This equation only one location). Showing bellow:

Thirsty= (thirsty _max- thirsty _value)*1/( thirsty _max- thirsty _min);
Weakness= (weak _max- weak _value)*1/( weak _max- weak _min);
OGT= (ogt _max- ogt _value)*1/( ogt _max- ogt _min);
Hbalc= (hbalc _max- hbalc _value)*1/( hbalc _max- hbalc _min);

*Table 3:* This score create the based on equation (This equation only one location)

| Kind of diabetes | Thirsty | Weakness | ……… | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 1 | 0 | ………. | 130 | 0 |
| Type2 | 0 | 0.2 | …….. | 150 | 0.5 |
| No diabetes | 0.5 | 1 | …….. | 135 | 1 |

*Step 4*

Next, weight adjustment each criteria particularly importance for client. After multiply weight and criterion score. Using this equation is showing bellow:

Thirsty= (thirst _max- thirst _value)*1/( thirst _max- thirst _min)*weight_ thirst;
Weakness= (weak _max- weak _value)*1/( weak _max- weak _min)*weight_ weak;
OGT= (ogt _max- ogt _value)*1/( ogt _max- ogt _min)*weight_ ogt;

*Table 4:* weight adjustment each criteria particularly importance for client

| Kinds of diabetes | Thirsty | Weakness | … | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 1*0.6=0.6 | 0*0.4=0 | … | 530*0.3=159 | 8 |
| Type2 | 0*0.6=0 | 0.2*0.4=0.08 | … | 450*0.3=135 | 6.5 |
| No diabetes | 0.5*0.6=0.3 | 1*0.4=0.4 | … | 135*0.3=40.5 | 5 |

*Step 5*

Finally, Add the all criteria value .Which totals are maximum this symptoms & test are selected. Using this equation is showing bellow:

Totals=Thirsty+ Weakness+……..+OGT+Hbalc;

*Table 5:* Totals are maximum this symptoms & test are selected

| Kinds of diabetes | Thirsty | Weakness | … | OGT | Hbalc | Totals |
|---|---|---|---|---|---|---|
| Type1 | 0.6 | 0 | … | 159 | 0 | 159.6 |
| Type2 | 0 | 0.08 | … | 135 | 0.05 | 135.13 |
| No diabetes | 0.3 | 0.4 | … | 40.5 | 0.1 | 41.3 |

a) *Total finally result*

*Table 6:* Total final result

| Kinds of diabetes | Criteria 1 | Criteria 2 | ………. | Criteria 11 | Criteria 12 |
|---|---|---|---|---|---|
| Type1 | Thirsty | Weakness | ……….. | OGT | Hbalc |
| Type2 | ‘’ | ‘’ | ………. | ‘’ | ‘’ |
| No diabetes | ‘’ | ‘’ | …….. | ‘’ | ‘’ |

| Kind of diabetes | Thirsty | Weakness | ……… | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 50 | 300 | ………. | 5500 | 2 |
| Type2 | 120 | 100 | …….. | 1500 | 3 |
| No diabetes | 100 | 50 | …….. | 2500 | 1 |

| Kind of diabetes | Thirsty | Weakness | ……… | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 1 | 0 | ………. | 0 | 0 |
| Type2 | 0 | 0.2 | …….. | 1 | 0.5 |
| No diabetes | 0.5 | 1 | …….. | 0.6 | 1 |

| Kind of diabetes | Thirsty | Weakness | …… | OGT | Hbalc |
|---|---|---|---|---|---|
| Type1 | 1*0.6=0.6 | 0*0.4=0 | …… | 0*0.3=0 | 2 |
| Type2 | 0*0.6=0 | 0.2*0.4=0.08 | …… | 1*0.3=0.3 | 8 |
| No diabetes | 0.5*0.6=.6 | 1*0.4=0.4 | …… | 0.6*0.3=0.18 | 1 |

| Kind of diabetes | Thirsty | Weakness | … | OGT | Hbalc | Totals |
|---|---|---|---|---|---|---|
| Type1 | 0.6 | 0 | … | 0 | 2 | 2.6 |
| Type2 | 0 | 0.08 | … | 0.3 | 8 | 8.43 |
| No diabetes | 0.3 | 0.4 | … | 0.18 | 1 | 1.98 |

Type2
Diabetes

Then final result provide. Look at the bellow.

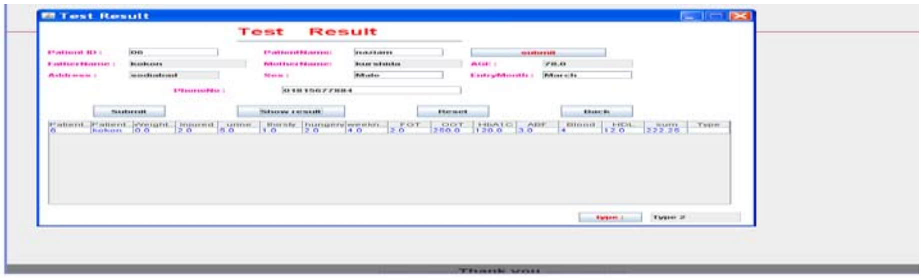



*Fig. 4:* Showing the test result Type2 diabetes

*Note:* one patient come at first entry, then symptoms selection then this system test provide then test selection then final test result provide this system provide type1 or type 2.

## V. Conclusion

The main goal of Computation Model for Identifying Types of Diabetics Using Multi-Selection Criteria Evaluation and K- Nearest Neighbor algorithm is to get best algorithms that describe given data from multiple aspects. There are different diabetic's symptoms classification algorithm that can be used for the identification of diabetes disease among patients. In this paper two classification techniques (MSCE, K-NN) are applied to predict the diabetes disease in patients. The algorithms are very necessary for intend an automatic classification tools. In our study first the two techniques were first filtered by using the computing time in which MCES helped the data set to be properly design and manipulated the system and K-Nearest Neighbor helped the diabetes range. The main purpose of applying this identification is diabetes level.

## References Références Referencias

1. The Utility of Feature Weighting in K-Nearest-Neighbor Algorithms - Ron Kohavi, PatLangley, Yeogirl Yun.
2. MCES (Multi-Selection Criteria Evaluation system) Algorithms.
3. Supervised Nurul Mustafa shaik shariful habib & Towhidul Islam "Computation Model for Identifying Types of Diabetics measurement system" Final paper or thesis. 23, January 2013.
4. Data collection & Analysis Dr. Hasan Ahmed. chittagond diabetes hospital limit.
5. Abdul Ghani, (1998), A. Medicinal plants of Bangladesh, PP 20-37, 54-56, Asiatic society of Bangladesh, Dhaka. Kirtikar and Basu, 1980.
6. Abdul Ghani, A. Medicinal plants of Bangladesh with chemical constituents and uses. 2nd edn. Asiatic society of Bangladesh.
7. Richard, 2011, the role of alternative medicine.
8. R.J. Vakil (Brit.Heart J., 1949, 11,350).