EXTRACTION OF KNOWLEDGE BY NEURONAL NETWORKS FROM DATA OVER PERSONS SUFFERING FROM DIABETES

Eugène MBUYI MUKENDI¹ Alpha MBUYI MITONGU² Richard KABAMBA ILUNGA²

1 Professor of the University of Kinshasa, Department of Computer Sciences, DR Congo

2 University of Kinshasa, Department of Computer Sciences, DR Congo

Laboratory of computer science in Kinshasa

Abstract: interesting results obtained by using neuronal network to extract knowledge from data on persons suffering from arterial high blood pressure [1] and [2] have encouraged us to use this method in other analysis of medical data.

We use neuronal networks from determining the importance of factors in the indication of complication of diabetes. Learning of multi-layered network is overseen and used an algorithm of retro propagation of gradient of error to estimate weights [1].

Keywords: learning, neuronal networks, algorithm, diabetes.

I. INTRODUCTION ^[1], ^[2], ^[6]

Diabetes is a disorder of assimilation, use and storage of

sugars brought by food. This result in a rate of glucose in the blood (called again blood sugar) high: we talk about hyperglaecemia. In spite of the medical research which advances every day, diabetes remains a disease that is treatable but does not cure. You need in your whole life, watch yourself, keep good eating habits and physical activity, be on drugs regularly. So a diabetic can be a patient in good health and at the quality of life.

More often than not the patient suffering from diabetes dies following the complications of the latter which may be cerebral, renal, due to the fact that it changes with time and patients do not know to follow up their treatment.

How to prevent these risks of complications and act on time from diagnosis taken over a patient? At present, there is no way allowing to evaluate automatically and accurately the risk that presents a patient to develop some complications of diabetes in the city of Kinshasa. So we suggest that at the end of this work to set up a computer science, decision-making system, which allows, from some factors gathered over a patient suffering from diabetes, to evaluate the risk that presents the latter to develop a complication by going by a set of data related to processed cases in the past.

To implement our application, we have modeled in UML and have used C- Sharp technology to implement our application. The sample on which we have had to work has been collected within a hospital complex of Mont-Amba located in the District of Lemba, within the University of Kinshasa between Pathological Neuro-Pshyco Center and the Private Hospital of the University of Kinshasa. It is during the period from October 10 to November 5 2012, which we had made up our sample.

I.1. Law of the learning

Let p and t be respectively input and output target vectors used for the learning of the perception and a the response of the perception. The progress of the value of w weight and means b = w oj of the perception is going to vary, every time input vectors are presented to the perception, under the following rule:

$$\Delta w = (t-a). p^t$$

Let e be be the number of epoch defined as follows:

$$e = t - a$$

then $\Delta W = (t - \alpha) = e$
And $\Delta b = (t - \alpha) = e$
So we will have: $W^{new} = W^{old} + \Delta W$
 $= W^{old} + e.p^t$
And $B^{new} = b^{old} + \Delta b = b^{old} + e$

I.2. Algorithm of the learning by learning by Widrow-Hoff

Let p and t be respectively input and output target vectors used for the learning of the network and a is the response of the network. The purpose of this algorithm is to minimize the average quadratic error between input and responses of the network.

Let F be the cost function, representing this error, defined as follows:

$$F = \frac{1}{n} \sum_{k=1}^{n} [t(k) - \alpha(k)]^2 = \frac{1}{n} \sum_{k=1}^{n} [e(k)]^2$$

Minimization of this function is done under the delta rule:

$$\Delta W = -\alpha \, \frac{\partial F}{\partial W}$$

The principle of the LMS algorithm (Least Mean Squared in English) is that it values at Kth iteration the average quadratic error by calculating the derivative of average quadratic errors with regard to the weight and by means of the network.

So:
$$\frac{\partial e^2(k)}{\partial w_j} = 2.e(k).\frac{\partial e(k)}{\partial w_j}$$
 et
 $\frac{\partial e^2(k)}{\partial b} = 2.e(k).\frac{\partial e(k)}{\partial b}$, j=1,2,...,R

Now
$$e(k) = t(k) - \alpha(k)$$

, $\alpha(k) = \sum_{i=1}^{r} w_i \cdot p_i(k) + b$]

This implies

$$\frac{\frac{\partial e(k)}{\partial w_j}}{\frac{\partial e(k)}{\partial b}} = \frac{\frac{\partial [t(k) - (w_p(k) - b)]}{\partial w_j}}{\frac{\partial e(k)}{\partial b}} \text{ and }$$

given that we have r input vectors of the perception, we obtain:

$$\frac{\partial s(k)}{\partial w_j} = \frac{\partial t(k)}{\partial w_j} - \frac{\partial [\sum_{i=1}^r w_i \cdot p_i(k) + b]}{\partial w_j} \quad \text{and}$$
$$\frac{\partial s(k)}{\partial b} = \frac{\partial t(k)}{\partial b} - \frac{\partial [\sum_{i=1}^r w_i \cdot p_i(k) + b]}{\partial b}]$$

By deriving partially with regard to W, and b, We simplify these two equations as follows:

$$\frac{\frac{\partial s(k)}{\partial w_j}}{\frac{\partial s(k)}{\partial b}} = -p_j(k) \qquad \text{and} \qquad$$

This means that weights and by means of the network must change respectively for 2ae (k) p (k) and 2ae (k) where aest the learning rate.

For the case of several neurones, we can generalize as follows:

W (k+1) = W (k) + $2ae(t).p^{T}(k)$ and b (k+1)= b(k) + 2ae(t)

What corresponds to the rule by Hebb

I.3. Back propagation or learning algorithm of «Back propagation»

Back propagation has been designed to generalize law of learning by Widrow-Hoff below for the multi-layered neuronal networks. Input vectors and corresponding target vectors are used to learn the network. Back propagation uses algorithm of descent of the gradient and attempts to reduce the difference between output of the network.

X pik and output wanted by applying the delta rule mentioned above.

The mistake made on Kth mode of the output is:

Input p of the neuronal network via the hidden neurone. Therefore, the total error (for all the modes) is :

To minimize this error, we calculate its gradient with regard to each w, weight, than we modify weights in the opposite direction of the gradient. Updating weights of the output layer. It is done as follows:

$$\nabla E_p = \frac{\partial E_p}{\partial W_{l,k,j}} = \frac{1}{2} \frac{\partial \sum_{k=1}^m (O_{pk} - x_{plk})^2}{\partial W_{l,k,j}}$$
$$\nabla E_p = -(O_{pk} - x_{plk}) \frac{\partial}{\partial W_{lkj}} [f(y_{plk})]$$

$$\nabla E_p = -(O_{pk} - x_{plk}) \frac{\partial f(y_{plk}) \partial (y_{plk})}{\partial w_{lkj} \partial (y_{plk})}$$

$$Or \begin{cases} \frac{\partial y_{plk}}{\partial w_{plk}} = \frac{\partial}{\partial w_{plk}} \sum_{j=0}^{m} (w_{lkj} - x_{p,i-1,j}) = x_{p,l-1,j} \\ \frac{\partial f(y_{plk})}{\partial y_{plk}} = f'(y_{plk}) \end{cases}$$

 $\nabla E_p = -(O_{pk} - x_{plk})f'(y_{plk})x_{p,l-1,j}$ $\nabla E_p = \delta_{pk}(1 - x_{plk})x_{p,l-1,j}$

The modification of weights varies according to calculation of the gradient. So, weights over the output layer are updated as follows:

Where u is the step to the learning between 0 and 1

1.4. Updating weights of hidden layers

To apply the technique below to the hidden layers, a priori we must know the wanted output of hidden modes. And yet, this is not the case for wanted output are not known.

To find solution to this problem, we must develop a term of error in the output of hidden nodes. We have no idea in advance what can be the correct or wanted output for these nodes. For that, we develop the term of the error in the output of the network as follows:

Xp. L-1.j depends on weights of the hidden layer and is given by: Xp1-1. J=fc

$$E_{p} = \frac{1}{2} \sum_{k=1}^{m} (O_{pk} - X_{plk})^{2} = \frac{1}{2} \sum_{k=1}^{m} [(O_{pk} - f(y_{plk})]^{2}$$
$$E_{p} = \frac{1}{2} \sum_{k=1}^{m} \left[O_{pk} - f(\sum_{j=0}^{n} W_{lkj} X_{p,l-1,j}) \right]^{2}$$

The gradient of Ep with regard to weights of hidden layers is worth:

The update on weights of the hidden layer is done in the opposite direction of the Ep gradient.

II. APPLICATION OF NEURONAL NETWORKS FOR DETERMINING THE IMPORTANCE OF FACTORS OF COMPLICATIONS OF DIABETES^[4],^[6]

Within the framework of this work, we are interested in four complications of diabetes that is say: diabetic retinopathy, diabetic nephropathy, infections and diabetic foot.

To make out some factors turned out to be important in the indication of complications of diabetes, we talked with many doctors and conducted documentary research.

Among these factors, let us give for example:

- Age
- Sex
- Tobacco
- Disease
- Non stable blood sugar (above 126)

- High blood pressure (consequence and factor of risk of risk)
- Eyes
- Medical history of ulcer or amputation
- Obesity
- Neuropathy (consequence and factor of risk of diabetes)
- Urea
- Creatinin
- Physical activity
- physical activity time

II.1. Presentation of the sample of the learning for method of neuronal networks

To constitute the sample of learning, we used data collected in the hospital complex of Mont Amba. These data have been stored in the relational data base access 2007 named BDT2. Database contains 3 tables: a patient table which contains identity of patient and a table diabetescomp which contains present factors in the complications of diabetes and different cases of complications of this disease.

Finally a table temps. To design our MCD, we have used rules such as: A patient can face zero or more cases of complications.

A case of complication concerns only one patient, because every time it happens a case, a patient has its own unique number that will be allocated.

II.2. Application of the method of neuronal networks

We have used SQL Server more accurately the console Microsoft Business Intelligence Development Studio for it integrates the modulus Analysis services which allows to carry out task of data mining.

The modulus analysis services contains many algorithms of data mining whose Microsoft neuronal network, which is the algorithm suggested by Microsoft for the method of neuronal networks.

TABLE OF IMPORTANCE OF FACTORS OVERCOMPLICATIONS OF DIABETES

Nom Facteur	Rétinopathie	Néphropathie	Infection	Pied diabétique
Age (31-54)	40,15	36,62	33,43	8,89
Sexe(m)	48,15	24,30	60,81	6,15
Antécédent ulcère	19,04	9,04	70,19	13,31
Alcool	61,29	3,22	51,41	6,60
Durée maladie (de 10–19)	70,56	22,22	26,54	11,93
Glycémie non corrigée	10,94	14,36	17,72	2,34
Hypertension	18,31	3,31	16,57	9,10
Yeux	13,83	3,23	1,75	1,69
Obésité	42,26	37,64	81,60	24,35
Neuropathie	20,73	11,29	34,06	5,34
Urée	30,70	70,24	21,60	11,71
Créatinine	41,99	56,13	20,47	17,81
Activitéphysique	16,43	16,99	39,66	15,19
Durée activité (2)	68,48	7,26	13,14	3,57
Durée activité (1)	11,84	20,74	38,59	4,42
Age (68-79)	14,45	10,14	65,78	7,72
Age (54-61)	11,79	19,62	48,94	7,41
Durée maladie (6- 10)	28,21	20,94	45,73	7,90
Durée maladie (3- 6)	14,14	20,99	54,65	6,87
Sexe (f)	47,10	22,02	14,77	8,33
Age (61-67)	19,29	14,23	57,53	7,39
Age (54-61)	11,79	19,62	48,94	7,41

II.3. Interpretation of results provided by the method of neuronal networks

Going from the application of the algorithm Microsoft Neuronal Network. To the data related to cases of complication of diabetes, results deduced from the importance of factors are the following: diabetics whose age varies from31 to 54 years old are more likely to develop retinopathy with 40% and nephropathy with 37% diabetic than other complications which is not the case for diabetics being in the gap from 68 to 79 years old who are more likely to develop diabetic infections than other complications.

Diabetics of male sex are exposed to infections. Diabetics of female sex carry the risk to suffer from retinopathy. Medical history of ulcer is an important factor leading to infections at diabetics with 70%. The medical history ulcer is a very important factor in the detection of infections.

- Alcohol is a very important factor which influences retinopathy with 81% and infections with 51%
- A diabetes who lasts more than 10 years leads to retinopathy with almost 71% and infections with 26%. The duration of disease proved as much useful to the detection of complications.

Diabetics having medical history of complications are exposed to a diabetic infection. Non corrected blood sugar (or non stable), causes at first infections then nephropathy and retinopathy. Arterial high blood pressure at a diabetic leads to the retinopathy and infections. Diabetics who have eyes problem (blurred vision, loss of vision, eyesight) never escape retinopathy.

Obesity is a very influential factor of all complications but more infections with 81%. Neuropathy, which is at the same time a factor of complication and a complication, is a factor leading to infections and retinopathy.

The positive urea influences with 70% nephropathy and retinopathy, it is very important. Diabetics whose creatinin is positive are the subject of nephropathy with 56% and retinopathy with 42% than other complications. Physical activity done during more than 30 minutes per diem allows to avoid complications whereas a patient who does physical activity in less than 30 minutes per week is subject to retinopathy.

III. DESIGN OF AN APPLICATION FOR THE APPLICATION FOR THE EVALUATION OF RISK OF COMPLICATIONS OF DIABETES

As for implementation, let us point out that database which has acted as example of learning has been designed by us.

Our contribution has consisted in, at first, designing an application of register of diabetics in the Democratic Republic of Congo, beginning by Kinshasa, fact which has not existed, what is the subject of our project, upon the request of HCMA in collaboration with WHO. The second contribution has consisted in developing a system of system of assistance to decision capable preventing patients from risks they carry to develop complications of diabetes.

Concerning our second contribution, we have designed a decision- making system which prevents the patient from risks he carries to develop complications of diabetes from some factors he presents,

This system is going to allow doctors to better take care of patients suffering from complications of diabetes.

The method of neuronal networks, has allowed to determine the importance of different factors by going by a sample of observed cases. So we can evaluate the risk that presents a patient with regard to one of complications all the more as already knows the importance of each factor. For this evaluation, we have set up an application in Java, capable of saving values of different factors and from this, evaluate the risk for each of our types of complications.

The table deduced from the application of the method of neuronal networks in this chapter gives probabilities of the indication of a complication when a factor is present. To evaluate the risk that presents a patient to show one of complication, we have questioned our model of exploration of data by DMX requests. That is to say that the model have made predictions, of course, upon our request.

III.1. Initial specifications of the application

We must design a system allowing the evaluation of the risk that presents a patient diabetes to develop one of four of complications. The system should allow the insertion of all new case of complication in data base, the evaluation of risks of complication, consultation of the chart of influence of factors along with chart of forecasting every number of the medical profession having access to data; webmaster will have the possibility all to use functionalities of the system.

III.2. Modeling

To model our system, we use the language of modeling UML2. Our system expanded according to the object-directed approach.

A. Diagram of class

After the analysis, has allowed us to present following classes:

- Patients
- Case of complication of diabetes
- Risks of complication of patients
- Factors
- Final users
- Webmaster

Associations related to different classes are the following:

- The webmaster manages data base by updating information (data) and addition of users.
- Patient develops a case of complication
- He presents risks of complication
- Factors allowing evaluation of risks of complication.
- Final user consults data base
- The case of complication gives information about a patient.

Hence the birth of our diagram of class:

III.3. DMX requests

DMX is a language close to SQL introduced by Microsoft for the handling of structures and models of exploration of data design under Microsoft.

DMX is used for prediction of values, in our case, we will use it to come to predict the risk of complications in order to bring doctors to better take care of their patients.

To lead to these predictions, we have made request under DMX and these are results which allowed us to bring details in the results that the system provides all the more as used the latter for the design of model and structure of exploration of data, for learning of data along with for prediction.

DMX requests that we have used for prediction are requests of singleton type for we have wanted to predict with regard to each factor, these requests are the following: For prediction of infections:

For retinopathy:

- SELECTFLATTENEDpredictHistogram(Infections_)asPneprhopathie FROM Comp Diabète1
- &"NATURAL PREDICTION JOIN" & "(SELECT'"& Activité &"'As Activité_Physique,'"
- & Ag &"'As Age,'" &Alcol& "'As Alcool,'" & Antécédent & "'As Antécédent Ulcère,'" &Créatinin& "' As Créatinine,'" & Durée & "As Durée Activité," & Durée & "Duréemaladie,'" "'As & Glycémie& Glycémie_non_Stable,'" & Hyper & "'As Hypertension,'" & Neuro & "'As Neuropathie,'" &obes& "'As Obésité,'" &sex& "'As Sexe,'" & Tab & "'As Tabac,'" &ur& "'As Urée,'" &yeu& "'As Yeux)As t

For retinopathy

- SELECT FLATTENED predictHistogram(Retinopathie_) as Pneprhopathie FROM Comp Diabète1
- &"NATURAL PREDICTION JOIN" & "(SELECT'''& Activité &"'As Activité_Physique,'''& Ag &"'As Age,''' &Alcol& "'As Alcool,''' & Antécédent & "'As Antécédent_Ulcère,''' &Créatinin& "' As Créatinine,''' & Durée & "'As Durée_Activité,''' & Durée & "'Duréemaladie,''' & Glycémie& "'As Glycémie_non_Stable,''' & Hyper & "'As Hypertension,''' & Neuro & "'As Neuropathie,''' &obes& "'As Obésité,'''

&sex& "'As Sexe,'" & Tab & "'As Tabac,'" &ur& "'As Urée,'" &yeu& "'As Yeux)As t

For nephropathy:

SELECT FLATTENED

predictHistogram(Infections_) as Pneprhopathie FROM Comp Diabète1 &"NATURAL PREDICTION JOIN" & "(SELECT'''& Activité &"'As Activité_Physique,'"

& Ag &"'As Age,'" &Alcol& "'As Alcool,'" & Antécédent & "'As Antécédent_Ulcère,'" &Créatinin& "' As Créatinine,'" & Durée & "'As Durée_Activité,'" & Durée & "'Duréemaladie,'" & Glycémie& "'As Glycémie_non_Stable,'" & Hyper & "'As Hypertension,'" & Neuro & "'As Neuropathie,'" & Neuro & "'As Neuropathie,'" & Sexe, "'As Obésité,'" &sex& "'As Sexe,'" & Tab & "'As Tabac,'" &ur& "'As Urée,'" &yeu& "'As Yeux)As t **Finally for diabetic foot :**

SELECT **FLATTENED** predictHistogram(Infections) as Pneprhopathie FROM Comp Diabète1 &"NATURAL PREDICTION JOIN" & "(SELECT'"& Activité &"'As Activité Physique,'" & Ag &"'As Age,'" & Alcol& "'As Alcool,'" & Antécédent & "'As Antécédent Ulcère,'" &Créatinin& "' As Créatinine,'" & Durée & "'As Durée Activité,'" & Durée & "'Duréemaladie," & Glycémie& "'As Glycémie_non_Stable," & Hyper & "'As Hypertension,'" & Neuro & "'As Neuropathie," & obes & "'As Obésité,'"

&sex& "'As Sexe,'" & Tab & "'As Tabac,'" &ur& "'As Urée,'" &yeu& "'As Yeux)As t

Some of captures of our application

<u></u>			
	Authentification		
Nom	Epah		
Code Accès	•••••		
Entant qu'Administrateur			

Figure 1 : authenticated form to have access to the application



Figure 2: recording form of patients in database.



Figure 3 : recording form of a new case of complication that appears in database.



157



Source Codes

package Diabète; //import com.jtattoo.plaf.mint.MintLookAndFeel; importjava.sql.ResultSet; importjava.sql.Statement; importjava.sql.Connection; importjava.sql.DriverManager; importjavax.swing.JOptionPane; /** * @author Horace

*/

public class Identité extendsjavax.swing.JFrame { Statementexecuteur = null; Connection conne; ResultSetresultat; Connection con; Statement stat; ResultSet res; /** Creates new form Identité */ publicIdentité() { try { UIManager.setLookAndFeel(new // MintLookAndFeel()); //pour le design } catch (Exception e) } //connexion bd try { Class.forName("sun.jdbc.odbc.JdbcOdbcDriver"); conne = DriverManager.getConnection("jdbc:odbc:Driver={Micro soft Access Driver (*.mdb)};DBQ= C:/BDT2.mdb"); executeur = conne.createStatement(); // JOptionPane.showMessageDialog(this,"Connectionreussi t"); System.out.println("Connection reussit"); } catch (Exception e) { System.out.println("echeque et mat"); }

initComponents();// pour le dimension et affichage au centre this.setLocationRelativeTo(null); this.setResizable(false); }private void BRechercherActionPerformed(java.awt.event.ActionEven tevt) { // TODO add your handling code here: try { // stat = con.createStatement(); resultat = executeur.executeQuery("Select * from Patient where IdPatient=""+txtrech+"""); // inti=0; while (resultat.next()) { TT.setValueAt(resultat.getString("IdPatient"), i, 0); TT.setValueAt(resultat.getString(2), i, 1); TT.setValueAt(resultat.getString(3), i, 2); TT.setValueAt(resultat.getString(4), i, 3); i++; } } catch (Exception e) { System.out.println("erreur de la connexion "); } } private void BEnregistrementActionPerformed(java.awt.event.Action Eventevt) { // TODO add your handling code here: try {int n = executeur.executeUpdate("INSERT INTO Patient VALUES("" + txtid.getText() + "","" + txtnom.getText() + "'," + txtpostnom.getText() + "'," + txtdate.getText() + "")"); if (n == 1) { JOptionPane.showMessageDialog(this, "ok"); } catch (Exception e) { JOptionPane.showMessageDialog(this, "erreur"+e); } private void BaffActionPerformed(java.awt.event.ActionEventevt) { // TODO add your handling code here: try { resultat = executeur.executeQuery("SELECT * FROM Patient"); inti = 0;while (resultat.next() == true) { TT.setValueAt(resultat.getString("IdPatient"), i, 0);

TT.setValueAt(resultat.getString(2), i, 1);

TT.setValueAt(resultat.getString(3), i, 2);

TT.setValueAt(resultat.getString(4), i, 3);

i++; } } catch (Exception e) { e.printStackTrace(); }

BSupActionPerformed(java.awt.event.ActionEventevt) {

// TODO add your handling code here:

 IJĊSI www.IJCSI.org

private void

```
supression();
private void
BAnnulerActionPerformed(java.awt.event.ActionEvente
vt) {
    // TODO add your handling code here:
txtnom.setText("");
txtpostnom.setText("");
txtdate.setText("");
txtid.setText("");
private void supression() {
int[] ligne = TT.getSelectedRows();
inti;
for (i = 0; i < ligne.length; i++) {
System.out.println(ligne[i]);
      String valeur =
String.valueOf(TT.getValueAt(ligne[i], 0));
try {
int res1 = executeur.executeUpdate("delete from Patient
where IdPatient = "' + txtid.getText() + """);
JOptionPane.showMessageDialog(this,
                                         "enregistrement
supprimer");
      } catch (Exception e) {
```

CONCLUSION

In order to help the medical profession take better care of patients, we have set up a decision –making system to come to their assistance in questioning.

By use of data mining, which brings together the set of techniques of extracting knowledge, more particularly neuronal networks, we have had to set up a system allowing the evaluation of risks of complication of diabetes at a patient.

The sample that we analyzed, come from the hospital complex of Mont Amba in the district of Lemba, within the University of Kinshasa, between Pathological neuropsycho Centre (PNPC) and the private hospital of the University of Kinshasa (PHUK).Going from this sample, we have devoted our study to four types of complication, that is to say: retinopathy, nephropathy, infections and diabetic foot.

REFERENCES

WORKS

1. MBUYI,E. et al.(2012), Datamining and Neuronal Net work II

DMX use for risk assessment of complication of arterial high blood, pressure, *in International Journal of Computer Science issues, IJCSI, volume 9, issue 5.*

- 2. MBUYI,E.et al(2012) Datamining and Neuronal Networks I. Extracting Kwowledge from high pressure data patients *in International Journal of Computer Science Issues,IJCSI, volume* 9, issue 3.
- 3. Manuel,D.et al (1999) *les systèmes décisionnels*, Conservatoire National des Arts et Métiers de Lille, Département Informatique Chaire Informatique :France
- 4. EKETE,F.(2011) Application du Datamining pour une meilleure pris en charge des malades atteint du diabète, bachelor's degree of science in computer science at the University of Kinshasa :DRC

- 5. LEBRATY, J.F. (2000), <u>les</u> systèmes décisionnels, IT Management Pinnaflex : France
- 6. KASENDE,I(2009) Application du Data Mining pour une Meilleure prise en charge des personnes atteintes d'hypertension Artérielle, Bachelor's degree of science in computer science at the University of Kinshasa, Faculty of sciences, Department of Mathematics and computer science .

