THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

USING SELF ORGANIZING MAPS NEURAL NETWORKS TO ANALYSE SYSTOLIC BLOOD PRESSURE BEHAVIOR BASED ON CLINICAL AND BIOCLINICAL PARAMETERS OF DIABETES PATIENTS

Alexandru DAIA[1], Ionela-Melania STANCIU[1], Nicoleta DRĂGANA[2,3] and Constantin Ionescu-TÎRGOVIŞTE[2,3]

[1] Faculty of Cybernetics Statistics and Economic Informatics, Bucharest, Romania
[2] "NC Paulescu" National Institute of Diabetes, Nutrition and Metabolic Diseases, Bucharest, Romania
[3] Clinic of Diabetes, Nutrition and Metabolic Diseases Medicine University "Carol Davila", Bucharest, Romania
Corresponding author: Alexandru DAIA, E-mail: alexandru130586@yandex.com

Accepted 13 March 2019

This article describes experimental results using Self Organizing Maps (SOM) also called Kohonen neural networks for analyzing systolic hypertension using clinical and bio clinical parameters. We have conducted experiments using Sompy library available as open source in Python on 900 diabetes patients and concluded this particular type of artificial neural networks could learn patterns regarding correlation of systolic hypertension with different other clinical or bio clinical values, which we consider surprising due to the fact self-organizing detects domain knowledge in data. The Sompy library is available at the following page: https://github.com/sevamoo/SOMPY. Our preliminary data shows that between various biochemical parameters in newly discovered diabetic patients there are interesting associations by clustering which can indicate some pathogenic mechanism and some clues for prevention and treatment.

Keywords: Diabetes, Neural Networks, Self-Organizing Maps, Machine Learning, Systolic blood pressure, Exploratory analysis.

INTRODUCTION

Both diabetes mellitus and hypertension are big and complex syndromes, each of them having several phenotypes. In addition, these two syndromes are often associated[1,2].

Our data shows that the type 2 diabetes, which is also associated with overweight or obesity, at their clinical onset, about 80% are already diagnosed as being treated for arterial hypertension[3].

One explanation could be that general practitioners easier diagnose blood pressure than the determination of blood glucose, which need a blood sample and a laboratory testing. In addition, a high blood pressure is associated to some clinical manifestations then a moderate high blood glucose level.[1, 2, 4, 5, 6, 7, 8, 23]

The mechanisms involved in these associations are complexes due to the heterogeneity of all metabolic changes associated with overweight /obesity, diabetes and dyslipidemia and also with high blood pressure.[9,10,11,12,13,14,15,16,17]

The aim of our investigation is that to see if, using a new artificial intelligence experimental approach, will results some unknown relationships between these two main distinct functions of the human body: hemodynamic and metabolic homeostasis: hypertension and various phenotypes of diabetes.

Self-Organizing Maps is named also Kohonen neural networks after the person who discovered them, Teuvo Kalevi Kohonen and emeritus professor at the Academy of Finland also a prominent Finnish academic and researcher according to Wikipedia.[18,19,20,21,22]

Self-Organizing Maps[21] provides a unique method for embedding a multidimensional space in a lower dimensional space named vector quantization in other words a data compression method with consideration that original information from the multidimensional space is preserved. This special type of neural network differs from other kinds of neural networks since it is a type of unsupervised method.

They have the ability to discover hidden nonlinear patterns in high dimensional data.

EXPERIMENTS AND RESULTS

As a graphical representation we could see in Figure 1 that self-organizing maps are represented by a planar grid of neurons, which represent the hidden layer in, which is fed the input layer in our case the subjects' parameters.

After the training procedure is finished, which is a heuristic similar with genetic algorithms, the yielded results help us visually to see also potential clusters.

For the experiment we have used recorded clinical and bio clinical parameters of 900 patients with diabetes mellitus Type 1 and diabetes mellitus Type 2.

In the plot bellow we can see the histograms of each clinical or bio clinical parameter. The right portion of the histograms corresponds to high values of parameters, which corresponds to the yellow cluster of neurons learned by the SOM as could be seen in the experimental results.

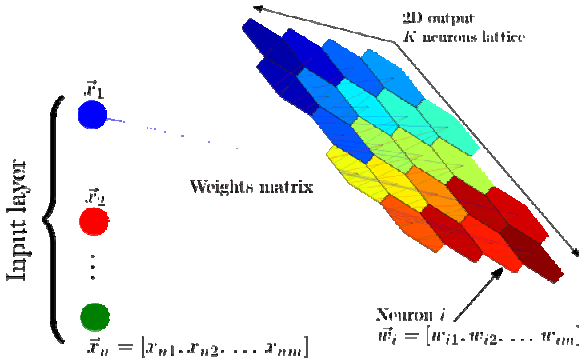

Figure 1. Overview of the SOM neural network.

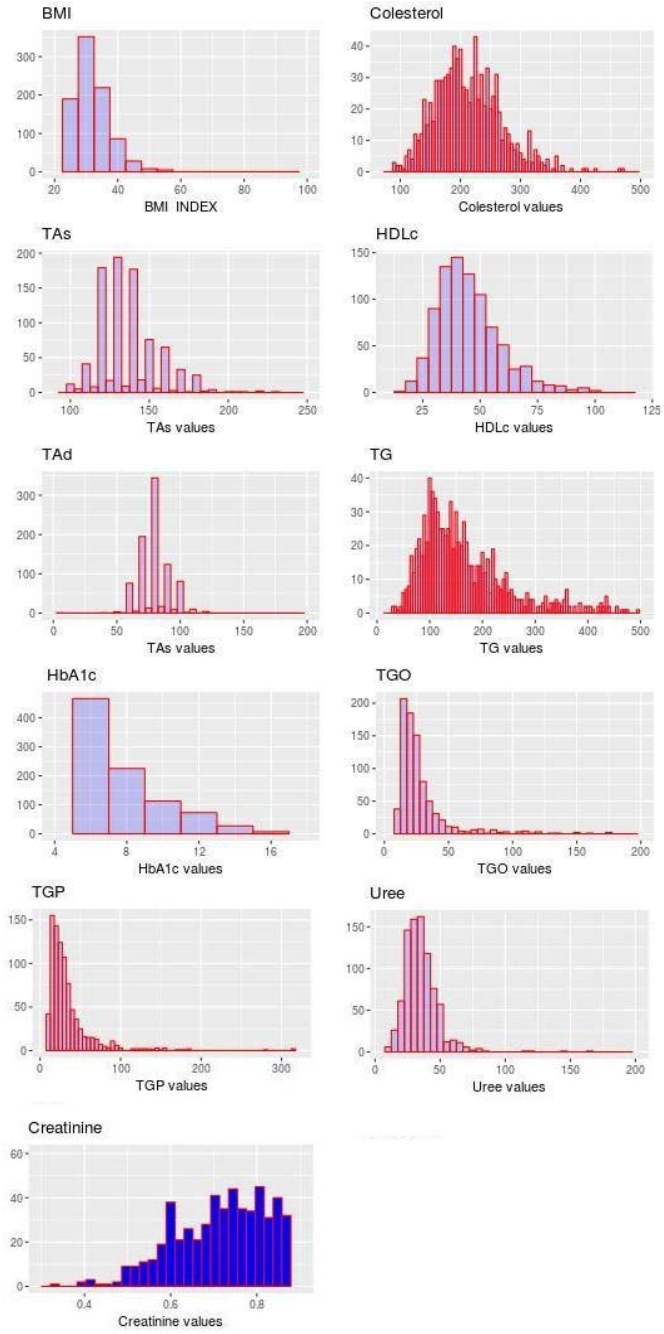

Figure 2. Histogram of clinical and bio clinical parameters. **BMI** – Body Mass Index [kg/m^2]; **TAs** – Systolic Blood Pressure, **Tad** – Diastolic Blood Pressure [mmHg], **HBA1c** – Glycated hemoglobin [%], Cholesterol [mg/dl]; **HDLc** – High Density Lipoprotein Cholesterol [mg/dl], **TG** – Triglyceride [mg/dl]; **TGO** – Glutamate-Oxaloacetate-Transaminase [UI/l]; **TGP** – Glutamate-Pyruvate Transaminase [UI/l]; Urea [mg/dl], Creatinine [mg/d].

The main interest in the experiment was to see behavior of systolic blood pressure according to the mentioned parameters.

Our expectations were focused on the Self-Organizing Maps: if they could detect patterns and hidden nonlinear correlation between the systolic blood pressure and the rest of parameters in order to see if the patterns yielded by the SOM are correct and what the patterns significance is, from a medical standpoint and considering the fact that the neural networks do not have a domain knowledge, but they could bring insights.

Our experimental phase is divided into two parts. First part was without Glycated Hemoglobin and the second one was with presence of Glycated Hemoglobin. In order to see different patterns and behaviors in the presence or absence of Glycated Hemoglobin hence, we considered this an influential factor since the patients from our data source where new patients when clinical and bio clinical measurements were made. So, the Glycated Hemoglobin reflects in some proportion the time the person had manifestations of presence of diabetes (both Type 1 and Type 2) before he was discovered as diabetes in hospital.

EXPERIMENT NUMBER 1
(without glycated hemoglobin)

Description

For this experiment, we used the data excepting glycated hemoglobin data as shown in following Figure:

	BMI	TAs	TAd	Colesterol	HDLc	LDLc	TG	TGO	TGP	Uree	Creatinina
75	29.555264	140.0	60.0	234.0	55.0	165.0	64.0	13.0	14.0	38.0	0.62
90	28.981143	150.0	90.0	325.0	40.0	243.0	213.0	29.0	13.0	22.0	0.78
138	28.936140	140.0	85.0	174.0	78.0	65.0	154.0	22.0	53.0	32.0	0.69
139	28.598552	220.0	100.0	262.0	54.0	191.0	87.0	25.0	21.0	43.0	0.89
143	28.875440	150.0	80.0	230.0	31.0	128.0	355.0	10.0	14.0	20.0	0.87

Figure 3. Sample of data from out cohort without HbA1c.

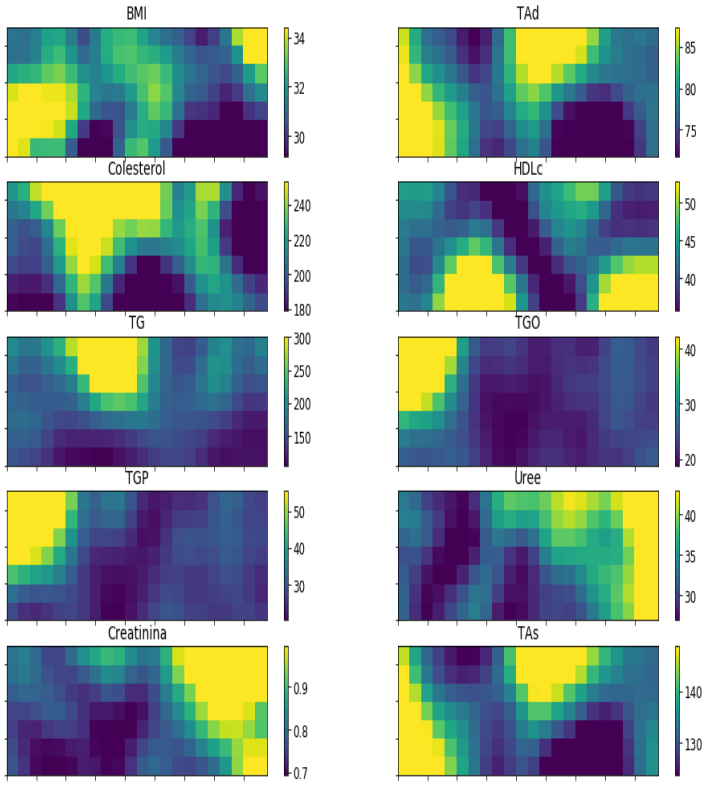

Figure 4. Component plane heat maps for each feature fed to the SOM without glycated hemoglobin, each “block” of colors representing a neuron.

After feeding and training the self-organizing map with the historical data contained in a csv (comma-separated values) file, without the Glycated Hemoglobin (HbA1c) from the table we obtained the following images presented in Figure 4.

Figure 4 in self-organizing maps terminology represents the heat maps of the component plane. The images there contains some grids, which could consider as clusters for which the self-organizing maps automatically calculates their sizes by calculating the eigenvalues of the data matrix during the training procedure. More explanation of the training procedure is beyond the scope on this paper. In other words, we need to consider that Figure 4 represents the Component plane heat maps for each feature feed into the training procedure (in our case the features represent the clinical or bio clinical parameters).

When exploring the Figure, we take in consideration that each heat map from Figure 4 represents the intensity of a particular features (parameter) learner by the self-organizing maps.

We also need to notice that when analyzing the figure each "block" or region of colors in each subfigure represents a unique neuron that is present in all other subfigures with similar or other shape but with same color.

Results for Experiment Number 1

- Nearly all big and very big values of systolic blood pressure (TAs) are present at:
- Persons for whom glutamate-pyruvate-transaminase (TGP) values were also high and very high.
- Persons for whom the level of creatinine was below the medium accepted values or very low values.
- Persons for whom the level of triglycerides (TG) was high and very high. In addition, there is a small percent of persons for whose level of triglycerides is low or normal, but the systolic blood pressure is high or very high.
- There are two big clusters of persons for whom the level of systolic blood pressure is high or very high. These clusters tend to exist and correlate with the cluster of people that exhibit high or very high values of high-density lipoprotein cholesterol (HDLc).
- High values of systolic blood pressure correspond with low values of Urea and low and medium values of systolic blood pressure corresponds with big values of Urea.
- From this, we can conclude that Urea and systolic blood pressure are ANTI-CORRELATED.
- Systolic blood pressure is nearly BUT **NOT PERFECTLY** correlated with diastolic blood pressure.
- The cluster of people for whom the high level of density lipoprotein cholesterol (HDLc) was very low is ANTI-CORRELATED with the cluster of people for whom the level of UREA was low.
- Glutamate-oxaloacetate-transaminase (TGO) and glutamate-pyruvate-transaminase (TGP) exhibit in general clusters of people for whom corresponding values are very small, small and medium, but in both exists a small cluster of high values for these parameters. In general, they are correlated.
- Creatinine exhibits a big cluster of people with high values, which is correlated with high values of high-density lipoprotein cholesterol (HDLc).
- There exist two big clusters of people with high values of systolic blood pressure (TAs) which are partially correlated with other two clusters of that exists in body mass index (BMI). We consider it partially correlated because
  one of the clusters with high body mass index is smaller meaning it contains fewer patients from which we can conclude that there exist persons with high values of systolic blood pressure but normal or small body mass index cases.

EXPERIMENT NUMBER 2
(with glycated hemoglobin)

Description

For this experiment, we fed into model glycated hemoglobin data as shown in following Figure 5.

The reason for feeding it is obvious, since it is a very important parameter for the fact that the patients were at first consultation at medical facility when medical examination was finished. It reflects how long time passed since the clinical manifestation of the disease become obvious and we concluded that would have a major impact on all other clinical and bio clinical parameters.

In Figure 6 are displayed the results after retraining the self-organizing map with the new parameter.

	BMI	TAs	TAd	HbA1c	Colesterol	HDLc	LDLc	TG	TGO	TGP	Uree	Creatinina
75	29.555264	140.0	60.0	6.0	234.0	55.0	165.0	64.0	13.0	14.0	38.0	0.62
90	28.981143	150.0	90.0	12.8	325.0	40.0	243.0	213.0	29.0	13.0	22.0	0.78
138	28.936140	140.0	85.0	6.0	174.0	78.0	65.0	154.0	22.0	53.0	32.0	0.69
139	28.598552	220.0	100.0	6.0	262.0	54.0	191.0	87.0	25.0	21.0	43.0	0.89
143	28.875440	150.0	80.0	6.8	230.0	31.0	128.0	355.0	10.0	14.0	20.0	0.87

Figure 5. Sample of data from out cohort with HbA1c.

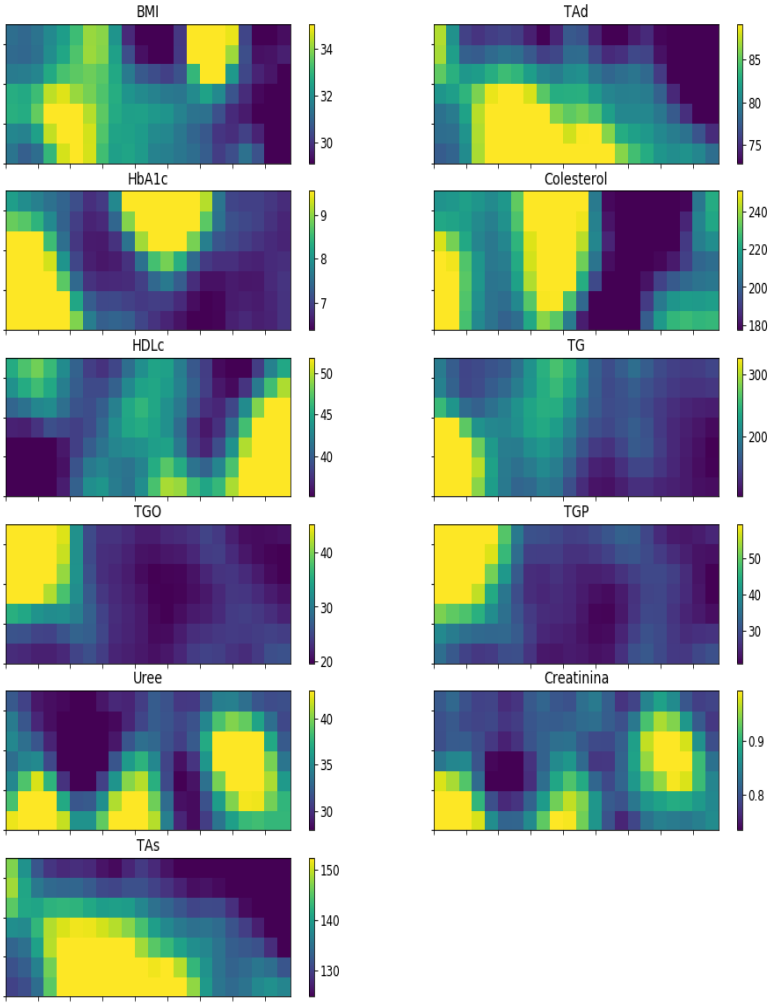

Figure 6. Component plane heat maps for each feature fed to the SOM with glycated hemoglobin, each "block" of colors representing a neuron.

Results for Experiment Number 2

As we see, the results have changed very much from first experiment.

Inspecting it, we could do the following observations, which are only several we consider the most important, but others could be derived especially if we increase the dimensionality of the data with new cases:

- The systolic blood tension has nearly only one big cluster with persons for whom the values are very high. This is encountered in high values of the HDLc.
- In a very interesting manner there are 3 clusters of persons who encountered very high values of urea and creatinine and more than that they look nearly perfect correlated.
- In urea, there are two clusters with very small values and in creatinine only one cluster.
- Body mass index (BMI) and glycated hemoglobin (HBA1c) contain two clusters of large values, which exhibit some degree of correlation.

- There are two nearly perfect correlated clusters of high values for glutamate-oxaloacetate-transaminase (TGO) and glutamate-pyruvate-transaminase (TGP). They are also a small proportion, the rest of the grid is dominated by medium, small and very small value clusters.
- Systolic blood pressure cluster with very high values is in this experiment correlated with high-density lipoprotein cholesterol (HDLc) and triglycerides.
- Systolic blood pressure high values cluster could be found at a smaller scale in one of the high values clusters of creatinine.
- Systolic blood pressure high values cluster is for some degree correlated with the glycated hemoglobin (HBA1c) high value cluster and since HBA1c exhibits also some correlation with body mass index BMI, results that Systolic blood pressure is at some degree directly correlated with glycated hemoglobin and by associativity rules correlated also with body mass index.

CONCLUSIONS

Our data shows that using a new method of prediction based on Self Organizing Maps (SOM) also called Kohonen neural networks can indicate the future potential complications in newly discovered diabetic patients, using current biochemical parameters.

REFERENCES

1. Ionescu-Tirgoviste C. Prolegomenon to the European Constitution Book of Diabetes Mellitus. Proc. Rom. Acad., Series B, **2008** *3*, 179–213, 2008.
2. Cooper ME, El-Osta A, Allen T J, *et al.*: Metabolic Karma-The Atherogenic Legacy of Diabetes: The 2017 Edwin Bierman Award Lecture. Diabetes **2018**, 67: 785790.
3. Ioacara S, Ionescu-Tirgoviste C.: Higher body mass index at diagnosis is a protective factor for mortality in a prospective trial of newly discovered type 2 diabetes patients, Diabetic Medicine 23, (Suppl. 4) P1654, 2006.
4. Ian H de Boer, Sripal Bangalore *et al.*: Diabetes and Hypertension: Position Statement by the American Diabetes Association. Diabetes Care **2017**, 40: 1273-1284.
5. Emdin Ca, Rahimi K, Neal B *et al.*: Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. JAMA **2015**, 313: 603-615.
6. Gregg EW, Li Y, Wang J *et al.* Changes in diabetes-related complications in the United States, 1990-2010. N Engl J Med 2014, 370:1514-1523.
7. Iliescu R, Lohmeier TE, Tudorancea I *et al.* Renal denervation for the treatment of resistant hypertension: review and clinical perspective. Am J Physiol Renal Physiol **2015**; 309: F583-F594.
8. Xie X, Atkins E, Lv J *et al.* Effects of intensive blood pressure lowering on cardiovascular and renal outcomes: updated systematic review and meta-analysis. Lancet 2016; 387:435-443.
9. Arauz-Pacheco C, Parrot MA, Raskin P. The treatment of hypertension in adults with diabetes. Diabetes Care **2002**, 25: 134-147.
10. Tong J, Boyko E.J, Utzschneider KM, *et al.*: Intra-abdominal fat accumulation predicts the development of the metabolic syndrome in non-diabetic Japanese-Americans. Diabetologia **2007**, 50:1156-1160.
11. National Cholesterol Education Program Expert Panel: ATP III final report: third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation and treatment of high blood cholesterol in adult's final report. I. Background and introduction. Circulation 2002, 106: 3157-3160.
12. Havel PJ: control of energy homeostasis and insulinaction by adipocyte hormones: leptin, acylation stimulating protein and adipolectin. Curr Opin Lipidol **2002**, 13:51-59.
13. Garancini MP, Calori G *et al.*: Prevalence of NIDDM and impaired glucose tolerance in Italy: a OGTT-based population study. Diabetologia **1995**, 38:306-313.
14. Perseghin G, Caumo A, Caloni M, Testolin G, Luzi L: Incorporation of the fasting plasma FFA concentration into QUICKI improves its association with insulin sensivity in nonobese individuals. J Clin Endocrinol Metab **2001**, 86: 4776-4781.
15. Hansson GK: Atherosclerosis:cell biology and lipoproteins. Curr Opin Lipidol **1998**, 9:73-75.
16. Targher G, Bertolini L, Scala L *et al.* Non-alcoholic hepatic steatosis and its relation to increased plasma biomarkers of inflammation and endothelial dysfunction in non-diabetic men. Role of visceral adipose tissue. Diabetic Medicine **2005**.
17. Piemonti L, Calori G, Mercalli A, Monti P, *et al.* Fasting Plasma Leptin, Tumor Necrosis Factor α Receptor 2 and Monocyte Chemoattracting Protein 1 Concentration in a Population of Glucose-Tolerant and Glucose-Intolerant Women – Impact on cardiovascular mortality. Diabetes Care **2003**, 26 (10) 2883-2889.
18. Hainan Lan, *Analyzing Climate Patterns with Self-Organizing Maps (SOMs),* https://towardsdatascience.com, **2018**.
19. Vahid Moosavi https://nbviewer.jupyter.org/gist/sevamoo/f1afe78af3cf6b8c4b67, **2014.**
20. https://www.niddk.nih.gov/health-information/diabetes/overview/tests-diagnosis
21. Kohonen Teuvo, Self-Organizing Maps, **2001**.
22. Klaus Obermayer, Terrence J Sejnowsky, *Self-Organizing Map Formation,*Foundations of Neural Computation, **2001**.
22. 1354-1358.
23. Arauz-Pacheco C, Parrot MA, Raskin P, *et al.* American Diabetes Association. Treatment of hypertension in adults with diabetes. Diabetes Care **2003**, 26: 580-582.