THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

DIABETES DETECTION WITH ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORK USING DERMATOGLIFES AS FEATURES

Alexandru DAIA [1], Nicoleta DRAGANA[2] , Janeta TUDOSOIU[2] and Constantin IONESCU-TÎRGOVIŞTE[2]

[1] Upwork Inc.
[2] "N.C. Paulescu" National Institute of Diabetes, Nutrition and Metabolic Diseases, Bucharest, Romania

Accepted November 28, 2019

The objective of this study is to create a proof of concept regarding using artificial intelligence to check existence of diabetes either of type 0 or type 1 from hand prints and to detect the risk of diabetes disease with automatic methods meaning ensemble of convolutional neural networks, with respect to the fact that the Holly Grail of diabetological science is early prevention. For this study right hand dermatoglyphes are used as feature for ensemble of convolutional neural networks.

Keywords: dermatoglyphes, diabetes, convolutional neural network, ensemble learning, majority vote, machine learning, prediction.

INTRODUCTION

The number of diabetes patients increased very fast in the last century and this disease became one of the first 10 causes of deaths worldwide after 2001, with an estimated global prevalence of 8.3% in 2013, which is expected to grow to an alarming figure of over 590 million people affected by 2035[1]. The highest percentage of population diagnosed with diabetes[2] is Marshall Islands (30.5%)[3] and lowest belongs to Benin (1.5%)[4]. The goal is to create a proof of concept to detect diabetes occurrence before the patients start to have the first symptoms.

More than 90% of all diabetics[5] suffer from Type 2 diabetes mellitus (T2DM). It is a multifactorial polygenic disorder that involves several endogenous and exogenous risk factors[6,7]. It is well known that type 2 diabetes is a disease with a well-defined genetic component, but also with a strong contribution of non-genetic components in its etiological complexity[8]. A number of prenatal environmental pathogens have been identified as possible risk factors for type 2 diabetes, including: mother's type of nutrition, objective difficulties or stress related to natural disasters, the season of birth (related to vitamin D metabolization) or melatonin), urbanization, drugs such as corticosteroids during pregnancy, pesticide use etc.[9–12].

One of the major goals of current diabetology is the prevention of both type 1 diabetes and type 2 diabetes[13]. The high prevalence of diabetes, as well as the upward evolution curve, is a feature of modern society, which can be appreciated by the accelerated growth of the GDP of the countries, starting with the middle of the twentieth century when the phenomenon was identified in the USA, later reaching the developed European countries, but also other countries such as Japan or Australia. After the Second World War, the division of Europe into two different zones, one considered as the capitalist zone in the west and the other as communist in the east, the incidence and prevalence of diabetes has experienced a much greater development in the western countries than in the eastern countries, a fact attributed to greater access to abundant food in the first group of countries compared to eastern countries. This finding led to the conclusion that besides the genetic factor known since the 19th century as being present in all populations, the second important factor, the one related to the lifestyle and especially the high availability of food, especially the high processed food, was considered the most important environmental factor.

This was well confirmed after 1990 when the western lifestyle was adopted by the eastern countries, with the increasing prevalence of diabetes. Type 1 and type 2 diabetes are two diseases that have different causes, but the predisposing genetic

substrate is involved in the pathogenesis of both. Only genetic changes are not capable of triggering the disease, triggers in the external environment play an essential role in the phenotypic manifestation of the disease for both types of diabetes. Type 1 diabetes, also known as juvenile diabetes, is an autoimmune disease that currently affects millions of people around the world, with an incidence of striking growth rates. According to the National Health Strategy 2014-2020, Romania ranks 9th in the EU as a prevalence of diabetes. Type 2 diabetes called "cancer of the 21st century" is a disease with a worrying increase in prevalence in the last half century. The western lifestyle characterized by lack of physical activity, high fat diet and high obesity rate led to this alarming change in genetic predisposition.

According to the European Association for the Study of Diabetes (EASD), approximately 195 million people suffer from diabetes worldwide and this number is expected to increase to 500 million by 2010.

The numerous dermatoglyphic models result from the association of a certain architecture of the many genes that encode the collagen molecules that are the basis of the generation of skin and plant skin relief accidents. A certain relief pattern is also identified in the skin of the other areas of the huge skin surface.

The unique character of dermatoglyphs, implies the possibility that among the many dermatoglyphic elements (models) that we will describe later, some of them may indicate the predisposition to different diseases. Leaving aside serious genetic diseases, which are usually monogenic defects and are associated with malformations or severe degenerative diseases that appear in childhood and which manifest dermatoglyphically through important and specific changes to each disease, expect us as in polygenic diseases as it is diabetes that may have some phenotypic characteristics of dermatoglyphs, able to predict to a greater or lesser extent the occurrence of diabetes. As all risk scores have advantages and disadvantages there are other information of this kind that can be used. Lately, there has been an increasing interest in the study of dermatoglyphs, as an indicator of genetic predispositions, including diabetes. Numerous researchers have published studies addressing this topic, the results indicating a clear correlation between diabetes and certain dermatoglyphic patterns. In 2016, these causal evidences between dermatoglyphic changes and diabetes both type 1 and type 2 materialized through a computer system for analyzing the number of digital ridges from the homologous fingers[14].

With the advancement of machine learning and artificial intelligence and specially algorithms as convolutional neuronal networks that are well established for image processing like face recognition algorithms, they just recently gain attention in medicine and especially pattern recognition in various fields of medical image processing.

This article proposes to explain existence of patterns regarding diabetes in hand dermatoglyphes using ensemble of Convolutional Neural Networks.

SHORT REVIEW OF LITERATURE

In 2018, Sehmi publishes a complex article on dermatoglyphes, which takes into account several studies published so far. The author's conclusion is that the dermatoglyphic investigation is absolutely cost-effective and does not require hospitalization and can help predict the phenotype of possible future diseases. Several studies conducted on different populations have identified a significant correlation between different patterns of fingerprinting and diabetes, however, the type identified varies from one geographical region to another, and this may be due to racial dermatoglyphic differences, which differ from one population to another. However, after a competent analysis, the results of the different authors seems to agree at a certain level. Possible reasons why not all study results correlate with each other could be, concludes Sehmi, small sample size, incomplete diagnoses, inadequacy of control group, statistical errors and lack of openness of subjects. Although the study of dermatoglyphs does not yet play a role in clinical diagnosis, it may serve as a way to select individuals from a larger population for further investigation, to subsequently confirm or exclude diabetes[15].

Another study, published in 2013, discuss the identification of quantitative parameters of dermatoglyphic analysis and qualitative parameters such as the model on the distal phalanges[16], as presented previously. They used handprints from 100 patients (50 men and 50 women) diagnosed with type insulin-independent diabetes (type 2 diabetes) and 100 normal fingerprints, from healthy persons (50 men and 50 women) as a control group. The patients were diagnosed by the doctor and were undergoing drug treatment. Subjects in the control group were selected on certain criteria, over 45 years of age, with no family history of diabetes and normal blood sugar level. Dermatoglyphs were read using a magnifying glass. In this paper four parameters were

studied, these include the dermatoglyphic models of the distal phalanx, angle atd, and line c.

In the fingers, the ulnar loops in the diabetic and non-diabetic were 2.60, respectively 5.89. This difference was statistically significant ($p < 0.05$). The whereabouts in diabetics and non-diabetics were 5.78 and 2.84 respectively. This difference was significant ($p < 0.05$). The radial loop and the arches showed no significant difference between diabetics and non-diabetics.

Many relevant studies discuss the quantitative aspects of dermatoglyphs. A larger angle atd and additional axial triads have been shown to be reliable indicators in the scientific screening of populations prone to diabetes. Rajnigandha *et al.* (2006)[17], Mittel and Lala (2013)[18], Sharma *et al.* (2013)[19]; Nayak *et al.* (2015)[20] showed an increase in angle atd in both sexes in diabetics while Mandascue *et al.* (2000)[21], Bala *et al.* (2015)[22], Srivastava and Rajasekhar (2014)[23] have shown that decreasing the angle atd would be specific for diabetics. Mandasescu *et al.* (2000)[21] showed that the atd angle of the right hand was significantly lower in male diabetics. Padmini *et al.* (2011)[24] showed an increase in angle atd only in male diabetics. Verbov *et al.* (1973)[25] showed a decrease in the number of a-b in female patients with type 1 diabetes. Similar findings were reported by Ziegler *et al.* (1993)[26].

SUBJECTS AND METHODS

The study included a total of 750 images with dermatoglyphs from both healthy and diabetes patients. The distributions consisted of 286 diabetes patients (both type1 and type 2 forms of diabetes) and 464 healthy persons. It can be observed that this is an unbalanced distribution of data.

Our method of obtaining dermatoglyphs was the classical ink method as can see in Figure 1.

The outcomes for this binary classification take were encoded with numbers in the set {0, 1} where:

- 0 means healthy patients
- 1 patient with diabetes.

METHOD

The method we used is named bootstrap aggregating, also known as bagging which represents an ensembling method[27]. The bagging method is well-known method used in machine learning by machine learning practitioners. It is in fact a meta-algorithm designed with the aims to improve robustness of machine learning models increasing performance in terms of accuracy of other performance metric in tasks of classification or regression.

The bagging method that we used works in the following way:

1. From a total of N images of recorded dermatoglyphics we split the entire dataset into 75% random images that will be considered training data, and the rest of remaining 25% images we store for evaluation using sampling with replacement
2. We train a convolutional neural network on the 75% of training data
3. We make predictions on the 25% and store the performances meaning accuracies
4. We repeat steps 1, 2, 3 for 51 times.
5. We aggregate the results meaning since we have accuracies obtained by the machine leaning model trained on different portions of the dataset 51 times and we proceeded to take the majority votes in terms of the majority class predicted by the model.

A general illustrative image of the bagging method can be seen in Figure 2.

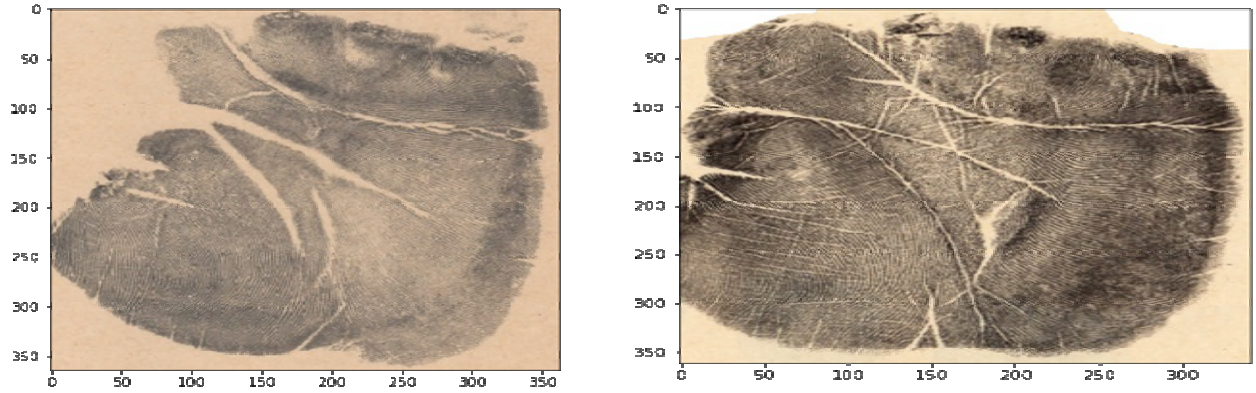

Figure 1. Palmar dermatoglyps captured through ink method.

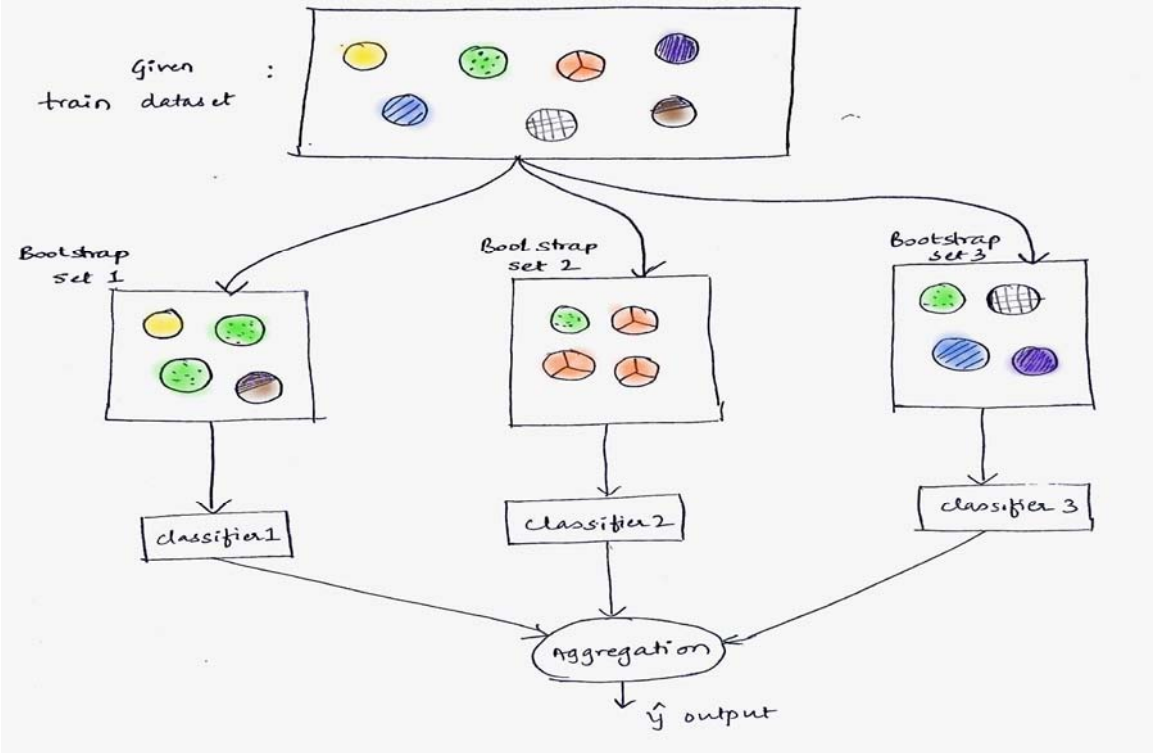

Figure 2. Illustrative schematic of bagging method[28].

Table 1

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	(None, 128, 128, 3)	0
zero_padding2d_9 (ZeroPaddin	(None, 134, 134, 3)	0
conv1 (Conv2D)	(None, 128, 128, 32)	4736
bn1 (BatchNormalization)	(None, 128, 128, 32)	128
activation_11 (Activation)	(None, 128, 128, 32)	0
dropout_17 (Dropout)	(None, 128, 128, 32)	0
zero_padding2d_10 (ZeroPaddi	(None, 134, 134, 32)	0
conv2 (Conv2D)	(None, 132, 132, 64)	18496
bn2 (BatchNormalization)	(None, 132, 132, 64)	256
activation_12 (Activation)	(None, 132, 132, 64)	0
dropout_18 (Dropout)	(None, 132, 132, 64)	0
max_pool2 (MaxPooling2D)	(None, 66, 66, 64)	0
zero_padding2d_11 (ZeroPaddi	(None, 70, 68, 64)	0
conv3 (Conv2D)	(None, 68, 66, 128)	73856
bn3 (BatchNormalization)	(None, 68, 66, 128)	512
activation_13 (Activation)	(None, 68, 66, 128)	0
dropout_19 (Dropout)	(None, 68, 66, 128)	0
max_pool3 (MaxPooling2D)	(None, 34, 33, 128)	0
conv4 (Conv2D)	(None, 32, 31, 64)	73792
bn4 (BatchNormalization)	(None, 32, 31, 64)	256
activation_14 (Activation)	(None, 32, 31, 64)	0
dropout_20 (Dropout)	(None, 32, 31, 64)	0
zero_padding2d_12 (ZeroPaddi	(None, 34, 33, 64)	0
conv5 (Conv2D)	(None, 32, 31, 32)	18464
bn5 (BatchNormalization)	(None, 32, 31, 32)	128
activation_15 (Activation)	(None, 32, 31, 32)	0
dropout_21 (Dropout)	(None, 32, 31, 32)	0
max_pool5 (MaxPooling2D)	(None, 10, 10, 32)	0
flatten_3 (Flatten)	(None, 3200)	0
fc1 (Dense)	(None, 256)	819456
dropout_22 (Dropout)	(None, 256)	0

Table 1 (continued)

fc2 (Dense)	(None, 128)	32896
dropout_23 (Dropout)	(None, 128)	0
fc3 (Dense)	(None, 128)	16512
dropout_24 (Dropout)	(None, 128)	0
fc4 (Dense)	(None, 2)	258

Total params: 1,059,746
Trainable params: 1,059,106
Non-trainable params: 640

The procedure involving steps from 1 to 5 was repeated 5 times with observation that we added some code such that in the majority vote we filtered out the neural networks that they tended to predict only diabetes class because of the unbalanced nature of the dataset, in this way we used only the models that performed well and we resolved the issue with unbalanced data.

The implementation was done in Python language, the most popular and rich language for machine learning and data science tasks.

The convolutional neural network model is Keras type, also a very popular neural network.

The images introduced in bagging ensemble model where reshaped in order to 128 width and 128 height in terms of pixels measurement.

Entire representation of our neural network could be seen in the following table with mention that for display simplicity we have not plot with shapes and layer names.

RESULTS AND CONCLUSIONS

The obtain accuracies are: 64%, 66%, 68%, 72%, 57%.

Since the variations of cross-validation results are high, this means we are still encountering over fitting. This is due to the fact that neural networks tend to work very well when data set is much larger. If in the future will enlarge the dataset size with more examples, let's say 5000, the standard deviation on the cross validation accuracies will decrease, making the model also more stable. Thus, our main goal was to have a proof of concept after experimenting this methods that concludes the fact that neural networks introduced in bagger ensemble capture an amount of information for classification between diabetes and non diabetes persons with the mentioned accuracies.

REFERENCES

1. International Diabetes Federation. IDF Diabetes Atlas (6th ed). ISBN: 2-930229. 2013; 32–49.
2. Anderson RN, Smith BL, "Deaths: leading causes for 2001 National Vital Stat Rep. 2003; 52:1–85.
3. C.L., Illsley. "Countries With The Highest Rates Of Diabetes." World Atlas, Oct. 1, 2018, worldatlas.com/articles/countries-with-the-highest-rates-of-diabetes.html.
4. Lynn, Jessica. "Countries With The Lowest Rates Of Diabetes." World Atlas, Apr. 25, 2017, worldatlas.com/articles/countries-with-the-lowest-rates-of-diabetes.html.
5. Rubino F: Is type 2 diabetes an operable intestinal disease? A provocative yet reasonable hypothesis. *Diabetes Care.* 2008; **31**(Suppl 2): S290–6.
6. Zia A, Kiani AK, Bhatti A, *et al.*: Genetic Susceptibility to Type 2 Diabetes and Implications for Therapy. *J Diabetes Metab.* 2013; **4**: 248–249.
7. Dayeh T, Volkov P, Salö S, *et al.*: Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 2014; **10**(3): e1004160.
8. Frayling TM: Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet.* 2007; **8**(9): 657–662.
9. Sobngwi E, Mbanya JC, Unwin NC, *et al.*: Exposure over the life course to an urban environment and its relation with obesity, diabetes, and hypertension in rural and urban Cameroon. *Int J Epidemiol.* 2004; **33**(4): 769–776.
10. Scott EM, Grant PJ: Neel revisited: the adipocyte, seasonality and type 2 diabetes. *Diabetologia.* 2006; **49**(7): 1462–66.
11. Dancause KN, Veru F, Andersen RE, *et al.*: Prenatal stress due to a natural disaster predicts insulin secretion in adolescence. *Early Hum Dev.* 2013; **89**(9): 773–6.
12. Greene NH, Pedersen LH, Liu S, *et al.*: Prenatal prescription corticosteroids and offspring diabetes: A national cohort study. *Int J Epidemiol.* 2013; **42**(1): 186–193.
13. Dragana N, Tudosoiu J, Daia A, Ionescu-Tîrgovişte C, "Dermatoglyphs and their importance for diabetes prediction. A review from the literature", Proceedings of the Romanian Academy – Series A: Mathematics, Physics, Technical Sciences, Information Science. 21, (2019).
14. Morris MR, Ludwar BC, Swingle E, *et al.*, A new method to assess asymmetry in fingerprints could be used as an early indicator of type 2 diabetes mellitus. J Diabetes Sci Technol, 10:864-871, 2016.
15. Sehmi S. Dermatoglyphic patterns in type 2 diabetes mellitus. Anatomy Journal of Africa, 7: 1162–1168, 2018.
16. Pathan JM, Rubeena N. Variations of Dermatoglyphic Features in Non Insulin Dependent Diabetes Mellitus. International Journal of Recent Trends in Science and Technology, 8:16-19, 2013.
17. Rajanigandha V. Mangala P, Latha P, Vasudha S.2006. Digito – Palmar complex in diabetes. Turk J Med Sci. 36:353-5.

18. Mittal M and Lala BS.2013. Dermatoglyphics: An economical tool for prediction of diabetes mellitus. Int J Med Health Sci. 2(3): 292-297.
19. Sharma MK and Sharma H.2012. Dermatoglyphics: A Diagnostic tool to predict diabetes. J of Clinic Diag Res. 6(3):327-32.
20. Nayak V, Shrivastava U, Kumar S, Balkund K.2015. Dermatoglyphic study of diabetes mellitus Type 2 in Maharashtrian population. Inter J Medical Sci Res Prac. 2(2):66-69.
21. Mandasescu S. Richards B, Cadman J.1999. Detection of pre- diabetics by palmar prints: a computer study leading to a low cost tool. XIV International Congress of the Federation for Medical Informatics. Dec 31; Germany. Manchester; GMDS: 2000.
22. Bala A, Deswal A, Sarmah PC, Khandalwal B, Tamang BK.2015. Palmar dermatoglyphic patterns in diabetes mellitus and diabetic with hypertension patients in Gangtok region .Int J Adv Res. 3(4):1117-25.
23. Rajnigandha V, Mangala P, Latha P, Vasudha S. Digito-palmar complex in NIDDM. Turk J Med Sci 2006:36(6):353-5.
24. Padmini MP, Rao BN, Malleswari.2011. The study of Dermatoglyphics in Diabetics of North Coastal Andhra Pradesh Population. I J Funda Appl Life Sci. 1(2):75-80.
25. Verbov JL.1973. Dermatoglyphics in early onset diabetes mellitus. Human Heredity. 23, 535542.
26. Ziegler AG, Mathies R, Mayer GZ, Baumgarti HJ, Rodewald A, Chopra V, Standl E.1993. Dermatoglyphics in type 1 diabetes mellitus. Diabet Med.10 (8):720-724.
27. Sabina Pokhrel, Beginners Guide to Convolutional Neural Networks, towardsdatascience.com/beginners-guide-to-understanding-convolutional-neural-networks-ae9ed58bb17d, 2019.
28. Hendrik Jacob van Veen, Le Nguyen The Dat, Armando Segnini. Kaggle Ensembling Guide. [accessed 2018 Feb 6]. https://mlwave.com/kaggle-ensembling-guide/, 2015.