

ANALYSIS OF THE PROCESS OF DEEP MACHINE LEARNING BASED ON THE RESULTS OBTAINED FOR PRIMARY DIAGNOSTICS OF GASTROENTEROLOGICAL DISEASES

**Rustam Yaxshiboyev¹, Bahodir Muminov², Erkin Ermetov³,
Abdunazar Jurayev⁴**

1 Information and Computer Technology and Programming
Tashkent University of Information Technologies named
after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan
E-mail: yaxshiboyevrustam@gmail.com

2 Software of Information Technologies Tashkent University of Information
Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan
E-mail: mbbahodir@tuit.uz

3 Department biomedical engineering, computer science and biophysics
Tashkent Medical Academy, Tashkent, Uzbekistan
E-mail: erkintashmi@yandex.ru

4 Samarkand State Veterinary Medicine, Livestock and Biotechnology
University Tashkent branch, Tashkent, Uzbekistan

ABSTRACT

This article discusses the analysis of the process of deep machine learning based on the results obtained for the primary diagnosis of gastroenterological diseases. The process of deep machine learning was carried out, the ANN and SVM algorithms were used. The results of deep machine learning were compared, two variants of data sets for training were collected. Based on deep machine learning, it is planned to further develop a software and hardware complex for the primary diagnosis of gastroenterological diseases.

Key words: deep machine learning, algorithm, forecasting, gastroenterological diseases, hardware and software complex, SVM, ANN.

INTRODUCTION

At present, the development of artificial intelligence in all countries of the world is developing carefully and rapidly. Links in the development of artificial

intelligence The President of the Republic of Uzbekistan Sh. Mirziyoyev issued a resolution “On measures to create conditions for the accelerated introduction of artificial intelligence technologies”, this resolution is in accordance with the strategy “Digital Uzbekistan - 2030” [1].

In the field of medicine, digital technologies can be widely used in the diagnosis, treatment of various diseases and different degrees. With the help of digital technologies, the work of doctors can be facilitated, the human factor is reduced, research time is reduced and efficiency is increased.

Within a short time, the doctor can make a decision about the diagnosis. With the help of digital technologies, controversial points can be overcome. Digital technologies use artificial intelligence, neural networks, machine learning and modern programming languages Python [8,9,10].

Artificial intelligence is the science and technology of creating intelligent machines, especially intelligent computer programs. AI is related to the similar task of using computers to understand human intelligence, but is not necessarily limited to biologically plausible methods.

In this process, the analysis of the process of deep machine learning based on the results obtained for the primary diagnosis of gastroenterological diseases is carried out.

The process of deep machine learning was carried out, the ANN and SVM algorithms were used. With the help of gastroenterologists, training materials were collected, that is, the data of the set from patients. Human saliva was chosen for primary diagnosis. The results of deep machine learning were compared. Based on these results, it is planned to develop a software and hardware complex called Saliva for the primary diagnosis of gastroenterological diseases.

Preliminary diagnostics helps to find out the problems, determine the bottlenecks of the enterprise, draw up a program for future changes and should answer the questions: is it possible to solve the identified problems, in what sequence they need to be solved [2,3,4].

In a scientific study, human saliva was taken. With the help of saliva, you can predict about gastroenterological diseases. With illness, the composition of saliva changes dramatically. The composition of saliva is the parameters [20].

By changing the composition of saliva, you can create a data set for training artificial intelligence algorithms. Table 1 shows the composition of a healthy person [14,15,16].

COMPOSITION OF THE SALIVA OF A HEALTHY PERSON

№	<i>The composition of saliva</i>	<i>Qty. (% and g/l)</i>
1	Water	99.4-99.5%
2	Organic and inorganic components	0.5-0.6%
3	Squirrels	1.4-6.4 g/l
4	Mucin	0.8-6.0 g/l
5	cholesterol	0.02-0.5 g/l
6	Glucose	0.1-0.3 g/l
7	Ammonium	0.01-0.12 g/l
8	Uric acid	0.005-0.03 g/l

Analysis of algorithms based on DEEP MACHINE LEARNING

In the article “Analysis of algorithms for predicting and preliminary diagnosis of gastroenterological diseases” [17,18,19], the initial results were taken, which contain 100 patients. As a continuation of this study, the number of patients was increased to 1000 per data set.

Patient parameters were obtained. Based on the obtained parameters, the data set was trained on the SVM and ANN algorithms.

The purpose of the algorithm involved in SVM:

In other words, “The goal is to maximize the minimum distance.” for the distance is given:

$$d_{H(\varphi(x_0))} = \frac{|w^T(\varphi(x_0)) + b|}{\|w\|_2}$$

$$w^* = \arg_w \max [\min_n d_H(\varphi(x_n))]$$

So, now that the goal is clear. By making predictions for the training data, which was binary, classified into positive and negative groups, if a point is replaced from the positive group in the hyperplane equation, we will get a value greater than 0 (zero), mathematically,

$$w^T(\varphi(x)) + b > 0$$

And predictions from the negative group in the hyperplane equation would give a negative value as

$$w^T(\Phi(x)) + b < 0.$$

But here the signs were about the training data, that is, how we train our model. This is for a positive class, give a positive sign, and for a negative class, give a negative sign.

But when testing this model on the test data, if we correctly predict a positive class (positive sign or sign greater than zero) as positive, then two positives results in a positive and therefore greater than zero result. The same applies if we correctly predict the negative group, since two negatives will again result in a positive result.

But if the model error classifies the positive group as negative, then one plus and one minus constitute a minus, hence less than zero overall.

To summarize the above concept:

The product of the predicted and the actual label will be greater than 0 (zero) if the prediction is correct, otherwise less than zero.

$$y_n[w^T \varphi(x) + b] = \begin{cases} \geq 0 & \text{if correct} \\ < 0 & \text{if incorrect} \end{cases}$$

For ideally separable datasets, the optimal hyperplane classifies all points correctly, additionally substituting the optimal values into the weight equation.

arg max is an abbreviation for max arguments, which are basically points in the area of the function where the function maximizes the values.

Also, taking the independent weight term outward gives:

$$w^* \arg_w \max \frac{1}{\|w\|_2} [\min_n y_n |w^T (\varphi(x) + b)|]$$

The inner term ($\min_n y_n |w^T \Phi(x) + b|$) basically represents the minimum distance from the point to the decision boundary and the nearest point to the decision boundary H.

Rescaling the distance to the nearest point as 1, i.e. ($\min_n y_n |w^T \Phi(x) + b| = 1$). Here the vectors remain in the same direction, and the hyperplane equation does not change. It's like zooming in on an image; objects expand or contract, but the directions stay the same and the image stays the same.

Distance rescaling is done by replacing:

$$w \rightarrow cw, \quad b \rightarrow cb$$

$$(cw)^T \varphi(x_n) + (cb) = c(w^T \varphi(x_n) + b) = 0$$

The equation now becomes (describing that each point is at least $1/\|w\|_2$ away from the hyperplane) as

$$w^* = \arg_w \max \frac{1}{\|w\|_2}, \text{ s. t. } \min_n y_n [w^T \varphi(x_n) + b] = 1$$

This maximization problem is equivalent to the following minimization problem, which is multiplied by a constant, since they do not affect the results.

An artificial neural network algorithm is a mathematical model, software or hardware implementation for organizing the construction of neural networks as a living organism. An artificial neural network is not programmed, but trained. Learning takes place in finding the coefficients of connections between neurons. The ability to learn is one of the advantages over algorithms.

The artificial neural network algorithm is also used in forecasting. This happens after training, that is, the neural network is able to predict the future value

of the sequence based on several previous values and those that exist at the moment.

Forecasting will be when the previous change in the actual degree overrides the future. For example, predicting a disease of a different type.

There are two types of learning process:

- Learning with a teacher
- Blended learning

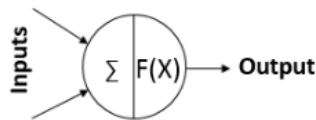


Fig1. Simple neuron

Activation of the input signal with the F(X) function. The activation function is sigmoid, ReLu, tanh, etc. This example uses a sigmoid activation function at layer nodes.

$$F(X) = \frac{1}{1 + e^{-x}}$$

Figure 1 shows a simple neuron. Now the task will be solved through the function. The value of the hidden layer is $1 = (1*0.1) + (1*0.1) + (1*0.1) = 0.3$.

$$Y_{in} = \sum X_i * W_{1ij}$$

Results

An example of the data of the set is indicated in table No. 2. The parameters and the name of the composition of human saliva are indicated.

An analysis of the ANN and SVM algorithms was made, the number of patients was 100 and 1000. The corresponding results were obtained [9,10,11].

Table 2

DATA SET PARAMETERS

<i>Data set parameters</i>	<i>The name of the composition of saliva</i>
Parameter_1	Protein
Parameter_2	Mucin
Parameter_3	cholesterol
Parameter_4	Glucose
Parameter_5	Ammonium
Parameter_6	Uric acid

The first time the training process was carried out on the number of 100 patients in the data set. (Figure 2,3)

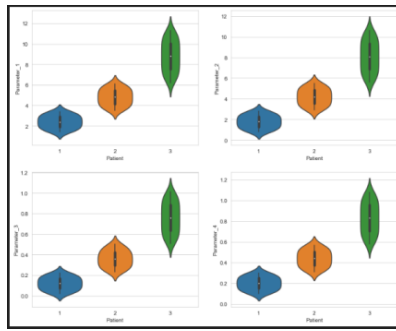


Fig 2. Determining the importance of parameters and predicting the likelihood of a patient's illness. (ANN)

In this algorithm, the result appeared on three colors. Accordingly, the result of training can be determined by colors:

- Blue - the probability of illness is higher
- Orange - the probability of illness is lower
- Green is healthy

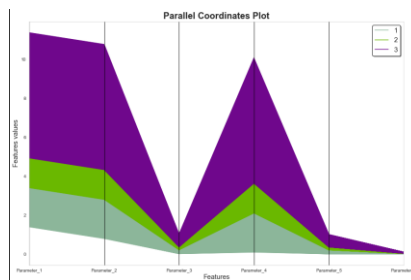


Fig 3. Determining the importance of parameters and predicting the likelihood of a patient's illness. (SVM)

In this algorithm, the result appeared on three colors. Accordingly, the result of training can be determined by colors:

- Violet - the probability of illness is higher
- Green - the probability of illness is lower
- Pistachio – healthy

Patient	Parameter_1	Parameter_2	Parameter_3	Parameter_4	Parameter_5	Parameter_6	
0	1	1.4	0.8	0.02	0.10	0.01	0.005
1	1	1.5	0.9	0.03	0.11	0.02	0.006
2	1	1.6	1.0	0.04	0.12	0.03	0.007
3	1	1.7	1.1	0.05	0.13	0.04	0.008
4	1	1.8	1.2	0.06	0.14	0.05	0.009
...
94	3	10.9	10.2	0.97	1.04	0.95	0.099
95	3	11.0	10.3	0.98	1.05	0.96	0.100
96	3	11.1	10.4	0.99	1.06	0.97	0.101
97	3	11.2	10.5	1.00	1.07	0.98	0.102
98	3	11.3	10.6	1.01	1.08	0.99	0.103

99 rows x 7 columns

Fig 3. Data set of 100 patients

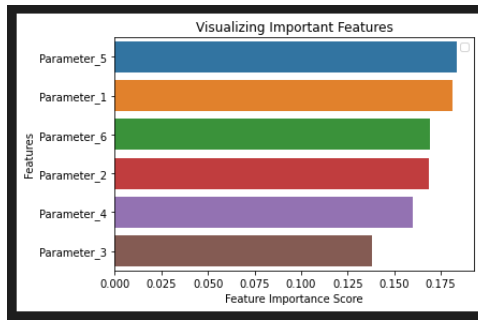


Fig 4. Importance of parameters from the data set (100)

The second time the training process was carried out on the number of 1000 patients in the data set.

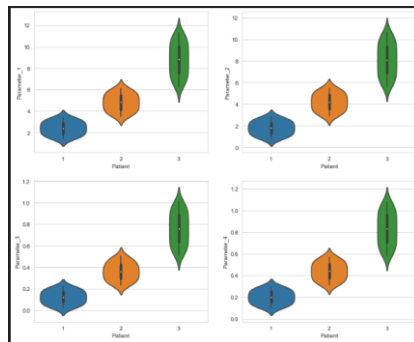


Fig 5. Determining the importance of parameters and predicting the likelihood of a patient's illness. (ANN)

In this algorithm, the result appeared on three colors. Accordingly, the result of training can be determined by colors:

- Blue - the probability of illness is higher
- Orange - the probability of illness is lower
- Green is healthy

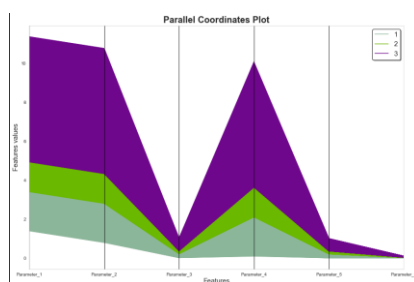


Fig 6. Determining the importance of parameters and predicting the likelihood of a patient's illness. (SVM)

In this algorithm, the result appeared on three colors. Accordingly, the result of training can be determined by colors:

- Violet - the probability of illness is higher
- Green - the probability of illness is lower
- Pistachio – healthy

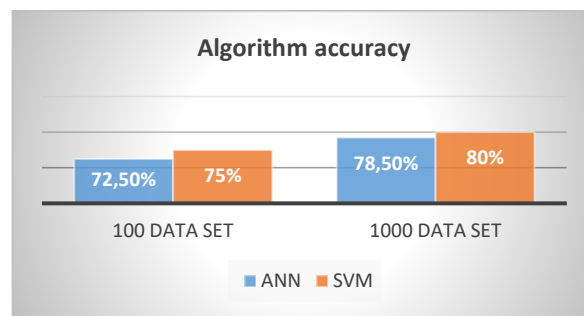


Fig 7. Algorithm accuracy between 100 and 1000 data sets

After deep machine learning of data sets, a comparison was made between the accuracy of ANN and SVM algorithms. Figure 7 showed the accuracy of the algorithms between 100 and 1000 data sets.

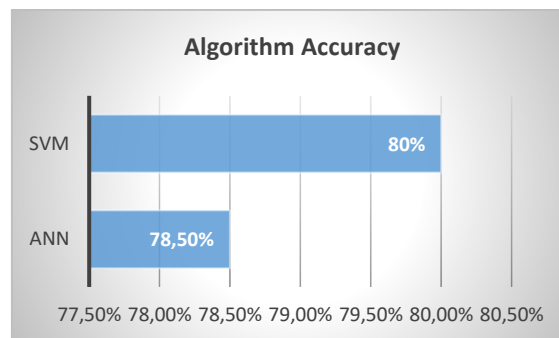


Fig 8. Algorithm accuracy between SVM and ANN

Figure 8 shows the overall accuracy of the algorithms after deep machine learning datasets. The overall accuracy of the algorithms allows scientific research to select a specific algorithm for the primary diagnosis of gastroenterological diseases.

Acknowledgment

The work was supported by the clinic of the Tashkent Medical Academy, the Department of Gastroenterology and the Department of Biomedical Engineering, Informatics and Biophysics.

CONCLUSION

As a result of the analysis and scientific research, an analysis of algorithms based on ANN and SVM deep machine learning was made. The data set was collected from 100 to 1000 patients. The accuracy of the algorithms between data sets of 100 and 1000 patients was 78.5% and 80%. Further scientific research will be carried out with other algorithms based on deep machine learning.

REFERENCES

1. <http://lex.uz//docs/5297051> О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта

2. Балашова А. Фейки и роботы: какими будут главные технологические тренды 2019.
3. Балашова А., Посыпкина А., Баленко Е. // РБК. – 2018. – 3 дек.
4. Christopher Bishop, Pattern Recognition and Machine Learning, 2006
5. Leon Stenneth, Philip S. Yu, Monitoring and mining GPS traces in transit space, SIAM International Conference on Data Mining
6. Ganesh J., Gupta M., Varma V. Interpretation of Semantic Tweet Representations // arXiv preprint arXiv:1704.00898. — 2017.
7. Zhang A., Culbertson B., Paritosh P. Characterizing Online Discussion Using Coarse Discourse Sequences // Proceedings of the International AAAI Conference on Web and Social Media. — 2017
8. Hastie, T., Tibshirani R., Friedman J. Chapter 15. Random Forests // The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.
9. Stallkamp J. et al. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition // Neural networks. - 2012. - Т. 32. - С. 323-332
10. Masci J. et al. Stacked convolutional auto-encoders for hierarchical feature extraction // Artificial Neural Networks and Machine Learning-ICANN 2011. - Springer Berlin Heidelberg, 2011. - С. 52-59.
11. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. - 2012. - С. 1097-1105
12. A nanoelectronics-blood-based diagnostic biomarker for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) <https://www.pnas.org/doi/full/10.1073/pnas.1901274116>
13. Микаелян Н.П., Комаров О.С., Давыдов В.В., Мейснер И.С. Биохимия ротовой жидкости в норме и при патологии. Учебно-методическое пособие для самостоятельной работы студентов по специальности «Стоматология». — Москва: ИКАР, 2017. — С. 10–11. — 64 с. — ISBN 978-5-7974-0574-0.
14. Физиология человека. Учебник. Под ред. В. М. Покровского, Г. Ф. Коротько. — М.: Медицина, 1997 ISBN 5-225-02693-1 т. 2 с. 39
15. Привес, М.Г. Анатомия человека / М.Г. Привес, Н.К. Лысенков, В.И. Бушкович. — Санкт-Петербург : СПбМАПО, 2011, 2014. — 720 с.
16. Анатомия человека. В 2 томах. Том 1 [Электронный ресурс]: учебник / Под ред. М.Р. Сапина - М. : ГЭОТАР-Медиа, 2013. - 528 с.

17. R. Nasimov, B. Muminov, S. Mirzahalilov and N. Nasimova, "Algorithm of Automatic Differentiation of Myocardial Infarction from Cardiomyopathy based on Electrocardiogram," 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT), 2020, pp. 1-5, doi: 10.1109/AICT50176.2020.9368738.

18. R. Nasimov, B. Muminov, S. Mirzahalilov and N. Nasimova, "A New Approach to Classifying Myocardial Infarction and Cardiomyopathy Using Deep Learning," 2020 International Conference on Information Science and Communications Technologies (ICISCT), 2020, pp. 1-5, doi: 10.1109/ICISCT50599.2020.9351386.

19. R. Yakhshibaev, B. Turaev, K. Jamolov, N. Atadjanova, E. Kim and N. Sayfullaeva, "Development of a mathematical model for balancing the level and device for remote monitoring of groundwater parameters," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670022.

20. Rustam Yaxshiboyev and Dilbar Yaxshiboyeva. "ANALYSIS OF ALGORITHMS FOR PREDICTION AND PRELIMINARY DIAGNOSTICS OF GASTROENTEROLOGICAL DISEASES." CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS) 1.2 (2022): 49-56