IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 4, Ser. II (Jul. –Aug. 2022), PP 39-43
www.iosrjournals.org

Oncodiagnosis Based on Blood Serum Microcalorimetric Data Using Machine Learning

Davit Khachidze[1], Papuna Karchava[2]
[1](E.Andronikashvili Institute of Physics, Iv.Javakhishvili Tbilisi State University, Georgia)
[2](Computer Sciences Department, Iv.Javakhishvili Tbilisi State University, Georgia)

Abstract:
Background*: The application of a machine learning approaches in medical diagnostics is becoming increasingly popular. The paper discusses machine learning as a task for classification of complex objects. Deep neural networks are commonly used in tasks for classification of complex objects. Medical diagnostics can be considered as a classification of complex systems. The blood serum and the data obtained on the basis of its microcalorimetric analysis are difficult objects. These data make it possible to diagnose onco-disease to a certain extent.*
Materials and Methods*: Deep neural networks are used for machine learning. The paper describes features of deep neural networks and analysis sigmoidal and logarithmic loss functions. By means of these functions, a model is defined, which allows to present microcalorimetric data of blood serum in a form usable for diagnosis. Data obtained by microcalorimetric analysis of blood plasma is used for a description of the system. The model created as a result of the use of machine learning, allow it to be used as a support for the pre-diagnosis of cancer. 300 records were used to test the performance of the model.*
Results*: The software implementation of the model is carried out by modifying the learning procedures of neural networks in the Python language library. It allows to read and process the data and predict the diagnosis based on the obtained results to support the doctor's decision.*
Key Word*: Machine Learning; Deep Neural Networks; Oncodiagnosis.*

Date of Submission: 10-08-2022 Date of Acceptance: 25-08-2022

I. Introduction

Any object, environment or process in the world can be perceived as a complex system. Many of these systems are variable in time and space, and it is often important to predict these changes and their effects. Medicine is considered as one such system.

The development of artificial intelligence, especially one of its directions, machine learning, plays a crucial role in our society. In many cases, it provides solutions to problems that were previously considered unsolvable. The development of technologies is accompanied by an increase in data flows, which in turn leads to a constant accumulation of data. This requires the use of more effective and large-scale methods for creating functional models in order to draw accurate conclusions.

One approach in machine learning is the training of deep neural networks [1]. It is based on a computational approach to neural networks that is popular in computer science. This model is seen as a new computing paradigm similar to the Von-Neumann or Turing machine. Over the past decade, many experiments have shown that neural "machines" can effectively solve many problems that standard sequential algorithms cannot. [2]. Therefore, neural networks have significant computational advantages that focus on model functionality.

We will discuss neural networks as a resource that can be used in decision support systems, in particularin medical diagnostics [3].

II. Supervised Machine Learning

There is a "generalized" definition of the process of machine learning: "Machine learning is the study of computer algorithms that allow computer programs to automatically improve with experience" [4, 5]. It is divided into two parts: first - supervised machine learning (Supervised ML) - the program learns with a large amount of data provided in advance, based on which it gives a correct answer to new data, and second - unsupervised machine learning (Unsupervised ML) the program is given a large amount of data, on the basis of which it must find a pattern for learning. Because of the fact that medical diagnosis is a complex system, supervised machine learning fits its description better [6].

The goal of supervised machine learning is to create a function *h(x)* that correctly predicts its value for the data *x* we want. As a rule, *x* is complex data.

Assuming that our target function is $h(x) = \theta_0 + \theta_1 x$, where θ_0 and θ_1 are constants, our goal is to find the best pair for prediction.

The predictor function *h(x)* is optimized using the learning material. For each learning sample we have input data *x* and its corresponding output *y*. Accordingly, for these samples we look for the difference between y and *h(x).* Using enough learning material, these differences allow us to estimate the error of *h(x).* After that, we change the pair θ_0 and θ_1 continuously until we get the best result. Based on these procedures, the predictor function *h(x)* is trained, and it can already make predictions for real data.

The so-called value function *J(θ)* is a measure of the well-chosen pair of θ_0 and θ_1, the most appropriate form of which for our task is as follows:

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{1}^{m}\left(h(x_{t,i}) - y\right)^2$$

According to the corresponding graph of this function (Fig. 1), our goal is to find such θ_0 and θ_1 that the value of the function is minimal, meaning it is located near the bottom. This is achieved by constantly changing this pair - periodically increasing one or decreasing the other, etc. This method is known as the gradient descent method, which is a well-known integral optimization algorithm. Since the main goal in machine learning is to maximize accuracy, the gradient descent method used to minimize errors is the most appropriate mechanism.

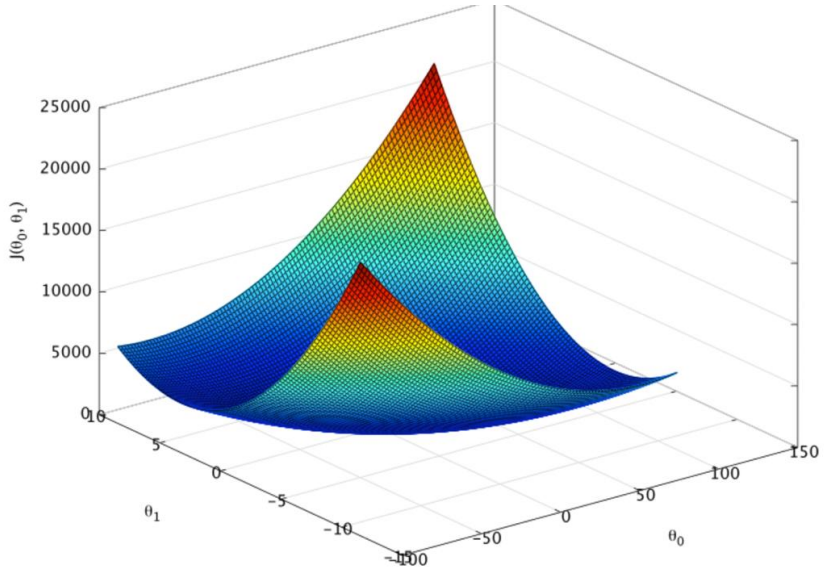

Fig.1 Graph of value function

When there is a large amount of data, multi-layer neural networks are used in machine learning. Each layer has input and output data with their corresponding weights and deviations. After passing each layer, the input value changes according to a specific function, and as a result of passing the last layer, we get the desired result. We use multi-layer, so-called deep neural networks, the number of layers of which is more than 3. Such networks are used for classification tasks, especially in diagnosis classification support systems in the field of medicine.

III. Microcalorimetric Data of Blood Serum for Diagnosis of Cancer Disease

The main object of our research, for which machine learning is used, is a method of diagnosing cancerous diseases based on differential scanning micro calorimeter (DSM) data. These are description of data by DSM on independent melting of major human serum proteins for both healthy and cancer patients [7, 8].

Based on many experiments, it has been established that the calorimetric curves of blood serum denaturation differ from the curves of non-oncological patients, and the curves of non-oncological patients differ from the curves of oncological patients. [9, 10] (Fig. 2.). Corresponding data for each curve for one patient is

not less than 5000. This amount of data in the presence of a large number of analyzes (experimental curves) creates a good basis for using machine learning.

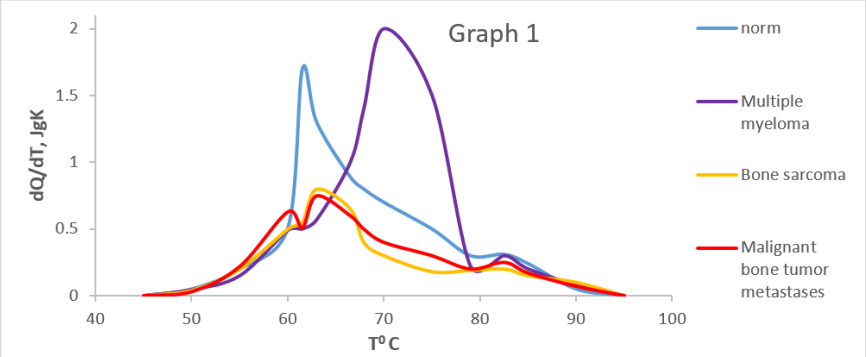

Fig. 2. Some microcalorimetric analysis curves of blood serum

IV. Learning Model - Development and Implementation

The learning model is presented in the form of a multi-layer neural network, which connects several layers together. Modern programming languages, such as Python, contain libraries that describe various commonly used algorithmic solutions. Among them are neural networks.

With this tool, a software package was created that reads the data obtained from the calorimetric analysis, defines the model, compiles the model, adjusts it, makes evaluations and "predictions".

Data reading - data is read from a text file in the format *x1;y1; x2;y2*, where *(x1, y1)* and *(x2, y2)* are points on the graph of the function. It is also possible to read data corresponding to several graphs at the same time.

Model definition - A neural network model consists of layers whose corresponding model is defined by the following code fragment:

```
model = Sequential()
model.add(Dense(12, input_dim=8, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```

The neural network of the model consists of three layers. The first and second layers are relu type. The first contains 12 neurons and has 8 parameters on the input, the second layer contains 8 neurons and the final layer contains 1 neuron to predict the result 0 or 1.

The Relu function is a detector function and is defined as follows:

$f(x) = max(0,x)$

This function gives good results in classification tasks.

The last layer uses a sigmoid function to make sure that the output is between 0 and 1.

Compiling the model - to compile the model, it is necessary to specify additional parameters that are required for learning the neural network. Learning a neural network means finding the best weights for our case. It is necessary to specify the same loss function, which in our case is a logarithmic loss function. We also use the optimized gradient descent algorithm adam to compile the model. As a result, we have the following code fragment,

```
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

running which compiles the model.

Model fitting - Model fitting refers to the process of learning a neural network:

```
model.fit(X, Y, epochs=150, batch_size=10)
```

The epochs parameter specifies the number of iterations.

Model evaluation - Model evaluation determines how well learning has occurred. At this stage, it is not evaluated what results the model will give us for real data, but the accuracy of the trained data is determined:

```
scores = model.evaluate(X, Y)
print("\n%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
```

In our case, the accuracy is 97%.

Prediction - The process of prediction involves determining a diagnosis based on data that did not participate in the learning process.

300 records were used to test the performance of the model. Testing was performed on different amounts of data. The smallest number corresponded to the data of 50 curves, the maximum - 300. The experiment showed that the accuracy increased with the increase of the data (table 1):

Table no 1. The result of experimental testing of the model.

Amount of data	Accuracy
50	69%
100	77%
150	84%
200	90%
250	94%
300	97%

By increasing the amount of learning data, the accuracy can be adjusted.

V. Conclusion

Decision support systems based on machine learning are becoming increasingly popular including in the field of medical diagnostics. The development and processing of various applications focused on the processing of large data flows, the source of which are various sensor (including medical) devices, is especially growing. The work carried out, the model created and its development potential will make it possible to equip a diagnostic tool – DSM - with a software application that will be able to support the decision to make a diagnosis with high accuracy based on the descriptive data of the blood serum denaturation curve.

References

[1]. Tom Mitchell, Machine Learning, McGraw Hill, 1997

[2]. N.Gabriel, N.F.Johnson Using Neural Architectures to Model Complex Dynamical Systems. Advances in Artificial Intelligence and Machine Learning. 2022;2(2):24.

[3]. Sergio Sanchez-Martinez , Oscar Camara, Gemma Piella, Maja Cikes, Miguel Ángel González-Ballester, Marius Miron, Alfredo Vellido, Emilia Gómez, Alan G. Fraser, Bart Bijnens. Machine Learning for Clinical Decision-Making: Challenges and Opportunities in Cardiovascular Imaging.. Frontiers in Cardiovascular Medicine. 04 January 2022. https://doi.org/10.3389/fcvm.2021.765693

[4]. Yu X, Efe MO, Kaynak O. A General Backpropagation Algorithm for Feedforward Neural Networks Learning. IEEE Transactions on Neural Networks. 2002;13:251–254.

[5]. McFall KS, Mahan JR. Artificial Neural Network Method for Solution of Boundary Value Problems With Exact Satisfaction of Arbitrary Boundary Conditions. IEEE Transactions on Neural Networks. 2009;20:1221–1233.

[6]. Imran Shafi, Sana Ansari,Sadia Din,Gwanggil Jeon, Anand Paul. Artificial neural networks as clinical decision support systems. Volume 33, Issue22 Special Issue: Applied Computational Intelligence (ACI 2019). Cognitive Computing for Robotic Vision (ISAIR2019). 25 November 2021. e6342

[7]. Khachidze D. and Monaselidze J. "Microcalorimetric study of human blood serum". Biophysics [IF SES 10(r)], vol. 45, №2, pp. 312-316, (2000).

[8]. D. G. Khachidze, N. D. Khachidze Microcalorimeter for diagnostics. International Scientific Conference Research Methods of Bio-Nano Agents Batumi, 24-26 November 2021

[9]. J. Monaselidze Ya. Kalandadze, D. Khachidze. "Microcalorimertic Investigation Of Influence Of Cancerogenic And Anticarcenogenic Compounds On Nuclear Chromatin Of Tumor Cells". Journal of Thermal Analysis, vol. 46, 431-440 (1996)

[10]. J. Monaselidze, M. Gorgoshidze, D. Khachidze, E. Lomidze, Sh. Barbakadze. Thermostability of chromatin and extracelular matrix in breast cancer tissue composition. International Conference, Physical Concepts of Nucleic-Acid Structure and Behavior, funded by VolkswagenStiftung, the state Committee of Science (Armenia) and ICTP network NET68, 27-29 May, 2013, Yerevan, Armenia, p.29.

[11]. J. Monaselidze, D. Khachidze, Z. Kuchadze, G. Tvauri, Sh. Barbakadze, I. Topuridze, M. Gadabadze. Thermal Stability of Blood Plasma Main Proteins of Cancer Patients. International Scientific Conference on Physical Research Methods in Medicine, Tbilisi, Proceedings I, pp. 181-183, 2011

[12]. J. Monaselidze, T. Lezhava, D. Khachidze, M. Gorgoshidze, M. Kiladze, E. Lomidze, G. Tvauri, Sh. Barbakadze, L. Kikaleishvili, M. Ramishvili, G. Nemsadze. The Possibilities of DSC in Breast Cancer Study and Diagnostics. International Scientific Conference on Physical Research Methods in Medicine, Tbilisi, Proceedings I, pp. 177-180, 2011

Davit Khachidze, et. al. "Oncodiagnosis Based on Blood Serum Microcalorimetric Data Using Machine Learning." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 24(4), 2022, pp. 39-43.