Classification of Coronavirus Disease with Artificial Intelligence and Machine Learning

Elif Akarsu [1*]

[1*] Ataturk University, Faculty of Engineering, Departmant of Electrical-Electronics Engineering, Erzurum, Turkey, (ORCID: 0000-0002-0338-3209), elif.akarsu@atauni.edu.tr

Abstract

Coronavirus disease is the most common contagious disease of recent times. The disease turned into a pandemic and spread throughout the world. The accuracy of the diagnosis of such a dangerous and contagious disease is of vital importance. In this study, a coronavirus diagnosis was made using artificial intelligence and machine learning, and the accuracy of the data obtained indicates how accurate the diagnosis is. The reason why the diagnosis of coronavirus was made based on the numerical results of other diseases in this study is to investigate the relationship of coronavirus disease with the findings of 15 different blood tests and to observe the positive and negative effects of this relationship on age. This study is important in this respect. 15 medical examinations were applied to 510 patients. The ages of the patients and the results of 15 examinations were recorded numerically. In addition, the coronavirus result of each patient was recorded. According to the patient's analysis results, the probability of being sick with coronavirus was obtained by comparing the real results with machine learning. As a result of the study, an accuracy rate of 89.6% was obtained.

Keywords: coronavirus, machine learning, artificial intelligence, deep learning, illness diagnosis.

Yapay Zeka ve Makine Öğrenmesi ile Koronavirüs Hastalığının Sınıflandırılması

Öz

Koronavirüs hastalığı son zamanların en sık görülen bulaşıcı hastalığıdır. Hastalık pandemiye dönüşmüş ve tüm dünyaya yayılmış durumda. Bu kadar tehlikeli ve bulaşıcı bir hastalığın teşhisinin doğruluğu hayati önem taşımaktadır. Bu çalışmada yapay zeka ve makine öğrenimi kullanılarak koronavirüs teşhisi konulmuştur. Ve elde edilen verilerin doğruluğu, teşhisin ne kadar doğru olduğunu ortaya koymaktadır. Bu çalışmada koronavirüs tanısının diğer hastalıkların sayısal sonuçlarına dayanılarak konulmasının nedeni, koronavirüs hastalığının 15 farklı kan testinin bulgularıyla ilişkisini araştırmak ve bu ilişkinin yaş üzerindeki olumlu ve olumsuz etkilerini gözlemlemektir. Bu çalışma bu açıdan oldukça önemlidir. 510 hastaya 15 sağlık muayenesi yapıldı. Hastaların yaşları ve 15 muayene sonuçları sayısal olarak kaydedildi. Ayrıca her hastanın koronavirüs sonucu kaydedildi. Hastanın analiz sonuçlarına göre, gerçek sonuçlar makine öğrenmesi ile karşılaştırılarak koronavirüse yakalanma olasılığı elde edildi. Çalışma sonucunda %89,6 doğruluk oranı elde edilmiştir.

Anahtar Kelimeler: Koronavirüs, Makine Öğrenmesi, Yapay Zeka, Derin Öğrenme, Hastalık teşhisi.

* Corresponding Author: elif.akarsu@atauni.edu.tr

1.Introduction

Machine Learning is the general name of computer algorithms that model a given problem according to the data obtained from the problem environment.[1] There are many proposed approaches and algorithms because it is a heavily studied subject. Some of these approaches are capable of prediction and estimation, and some of them are capable of classification. There are many recommended machine learning methods available. These may differ in approaches to the problem and therefore may have different successes in different problems.

Machine learning is a subset of artificial intelligence.[2] An "intelligent" computer thinks like a human and performs tasks on its own. One way to train a computer to mimic human thinking is to use a neural network, a set of algorithms modeled on the human brain. Machine learning is a type of predictive analysis. But it is much easier to implement machine learning, which is updated in real time as we receive more data.

Machine learning, unlike other artificial intelligence applications, is a set of algorithms that do not need rules that we can interpret and enter manually, while imitating human intelligence on the one hand. Just as a person learns concepts by himself/herself with what he/she sees and hears and does not need someone to process these rules in his/her brain, machine learning applications similarly learn the desired task by assimilating the datasets presented to him/her. We can explain the difference between artificial intelligence algorithms that do not contain machine learning, which we can call "classical" or "traditional" algorithms, and machine learning algorithms with the following simple example: Imagine a human baby who has no knowledge of animals. This baby will not be able to distinguish between cats and dogs. In classical artificial intelligence applications, instead of teaching the baby the difference between cat and dog, we give him a full set of instructions that will enable him to distinguish between cat and dog. These instructions should be prepared in such a way that this ignorant baby should be able to both follow those directions and distinguish between cat and dog mostly correctly. It is obvious how difficult this is for both the baby and the teacher. On the other hand, as you may have noticed, we don't really need to do this to increase baby's success: If we give baby lots of pictures of cats and dogs and repeatedly tell which is which, after a point the baby will begin to distinguish the two with high accuracy. This is exactly how machine learning algorithms work, as the name suggests. Of course, computers unfortunately do not come with a learning algorithm in them when they are built. Computer scientists themselves design and adapt the most suitable machine learning algorithm for each problem. Machine learning is essentially built around this logic. Many artificial intelligence applications that we have seen recently are actually machine learning applications. Machine learning is indispensable in real life for most artificial intelligence applications that we are used to seeing in science fiction movies; Therefore, the concepts of "machine learning" and "artificial intelligence" are getting more confused. At this point, although we have largely avoided the conceptual confusion by revealing the difference between the concepts of machine learning and artificial intelligence, there is one more step we need to take: Although machine learning is a narrower concept than artificial intelligence, it still includes many different algorithms and algorithms. It is a method field. That's why the concept of "machine learning" still doesn't adequately describe the products of popularization we're talking about.

Types of Machine Learning Algorithms

Machine learning algorithms are divided into 4 categories, each designed for a different purpose. For example, supervised learning is for scaling the scope of data and making predictions based on it. On the other hand, unsupervised algorithms are used to organize and filter data so that it makes sense.

Supervised Machine Learning

Algorithms that require some supervision of the developer during the process are known as supervised machine learning. The developer tags the training data and sets the strict rules and limits that the algorithm will follow. Supervised machine learning algorithms can apply what has been learned in the past to new data, using labeled examples to predict future events.[3] In supervised machine learning, the goal is to predict the target variable using a function defined over a set of arguments. Audited algorithms work by identifying a set of input data and expected results. An algorithm is considered successful when its matches and predictions are found to be correct. The learning algorithm can also compare its output with the correct output and find errors in order to modify the model accordingly.[4] One of the most famous examples of supervised machine learning is the Boston housing prices dataset. This dataset, which includes houses sold, their properties, and their selling prices, aimed to create a machine learning model that can predict the selling price of any house.

Unsupervised Machine Learning

Unsupervised learning is the approach we use when we have little or no idea what the output we want from our data will look like. We can create the model from the data that we do not know the effect of the variables. In "unsupervised learning" there are only data, no information is given about them. Attempts are made to draw conclusions from these data. Since no information is given about the data from the beginning, it cannot be said that the conclusions drawn are absolutely correct.

Semi-Supervised Machine Learning

Algorithms that combine features of both supervised and unsupervised algorithms are defined as semi-supervised machine learning. Not all of the training data is labeled and not all rules are provided when starting the algorithm. Semi-supervised machine learning algorithms use both labeled and unlabeled data for training. Systems that typically use small amounts of labeled data and large amounts of unlabeled data can significantly improve learning accuracy.

2. Material and Method

15 medical examinations were applied to 510 patients. The ages of the patients and the results of 15 examinations were recorded numerically. In addition, the coronavirus result of each patient was recorded. According to the patient's analysis results, the probability of being sick with coronavirus was obtained by comparing the real results with machine learning. As a result of the study, an accuracy rate of 89.6% was obtained.

Table 1. 15 types of Diseases

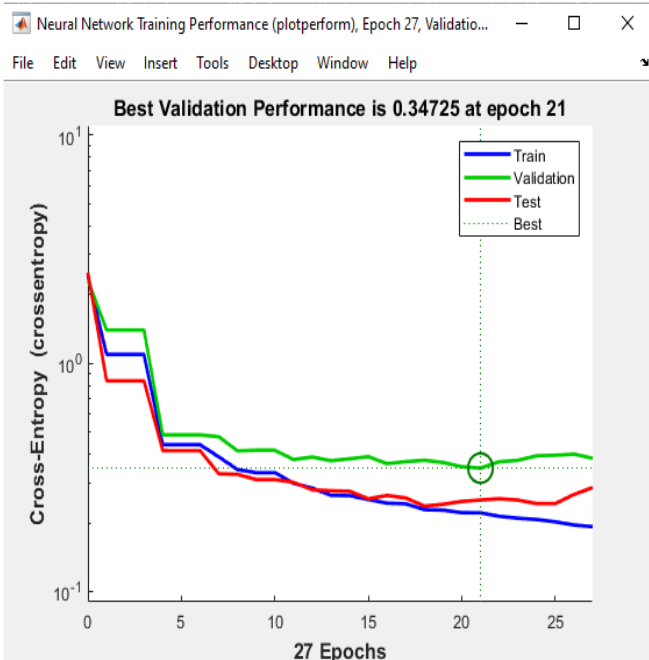| |
|---|
| 'Hematocrit' |
| 'Hemoglobin' |
| 'Platelets' |
| 'Mean platelet volume ' |
| 'Red blood Cells' |
| 'Lymphocytes' |
| 'Mean corpuscular hemoglobin concentration (MCHC)' |
| 'Leukocytes' |
| 'Basophils' |
| 'Mean corpuscular hemoglobin (MCH)' |
| 'Eosinophils' |
| 'Mean corpuscular volume (MCV)' |
| 'Monocytes' |
| 'Red blood cell distribution width (RDW)' |
| 'Neutrophils' |

Figure 1. Cross-Entropy

The data set was reviewed 21 times, and as a result of this process, the learning process was completed, and as a result, the iteration stopped. As a result of all iterations, the mean square error was obtained as 0.34725. The smaller this value, the better the system performs.
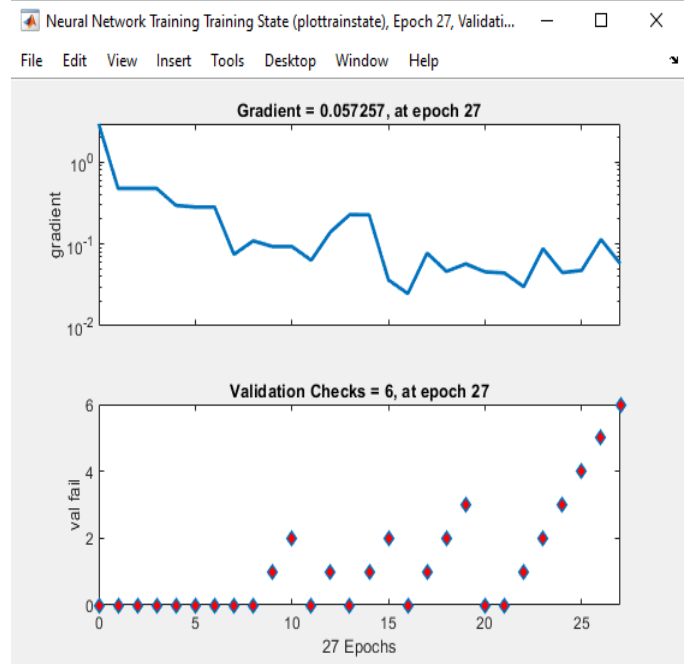

Figure 3. Training State

The learning process continues until 6 validation checks are completed. average gradient value is 0.057257.
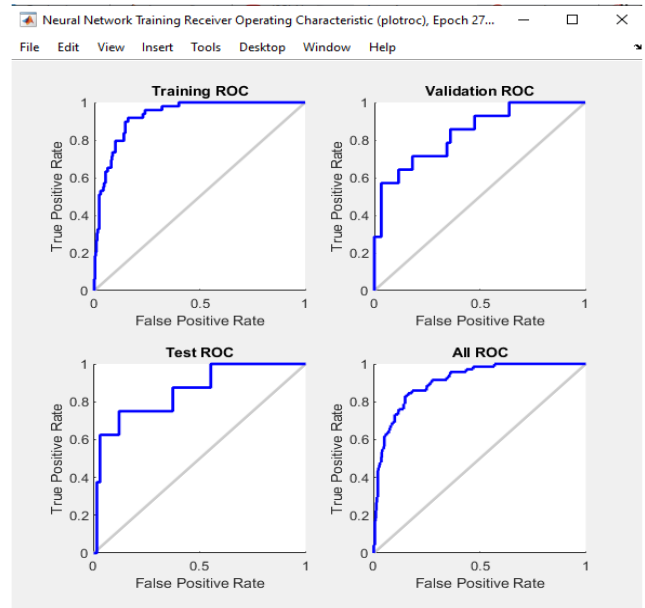

Figure 2. Receiver Operating Characteristic

Figure 2 shows ROC graphs. The 1st graph shows the training. The closer this graph is to a value of 1, the better. If it goes to the side drawn in red on the upper left side of the other 3 graphs, it is said that the learning was so successful. The average of all forecasts is plotted on the All ROC graph.
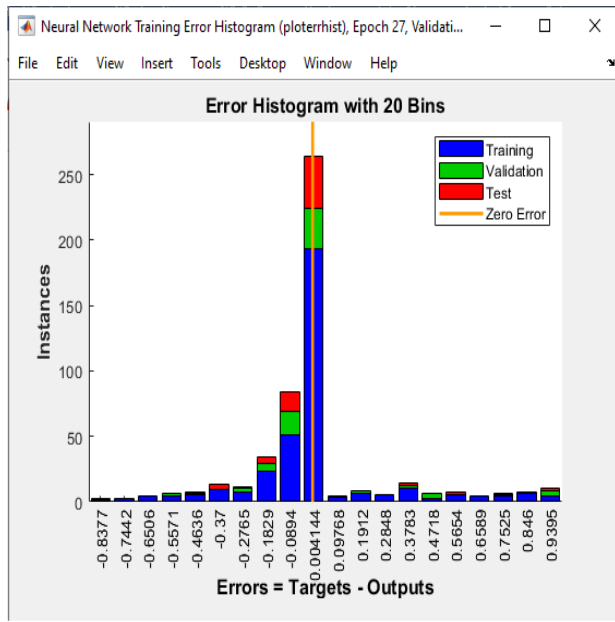
Figure 4. Error Histogram

It is seen that the highest error is between 0.004144 and -0.0694. This is also a very small value.

3. Results and Discussion

In the 1st matrix, it is seen that the education is successful at the rate of 89.9%. Against a total of 510 inputs, the results were predicted and highly accurate predictions were made. An accuracy of 85.3% was achieved in the validation confusion matrix. With this matrix, the accuracy of the test data was determined. An accuracy of 92% is seen in the test matrix against the training done. It's a pretty good rate. The average accuracy of all confusion matrices was 89.6%. It was aimed to investigate the relationship between coronavirus disease and the imbalance in blood cells, and it was aimed to reveal the harmony between respiratory tract syndrome and blood values. And the accuracy of the diagnosis of such a dangerous and contagious disease is of vital importance. The numerical result obtained shows that a significant degree of accuracy has been achieved. This shows that in the future, it will be investigated not only as a respiratory tract infection, but also as a disease that will be associated with different diseases.
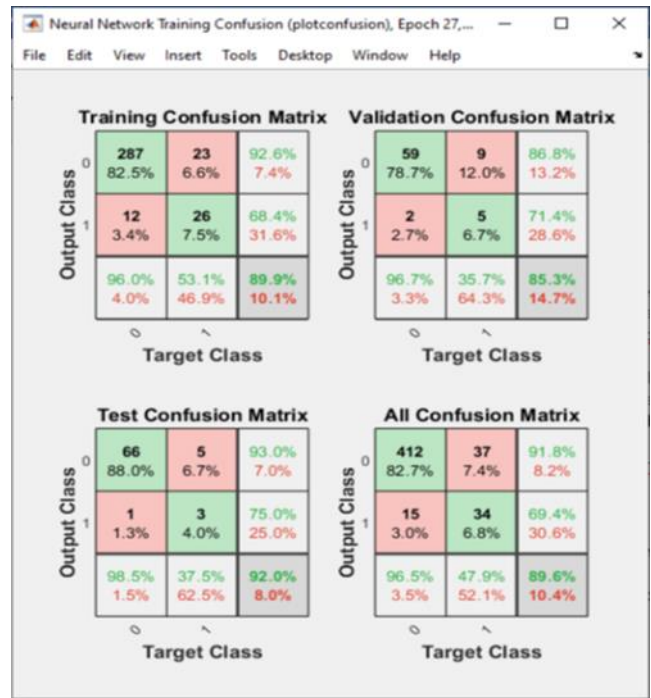

Figure 5. Confusion Matrix

4. Conclusions and Recommendations

In this study, the possibility of carrying the coronavirus was examined based on the data obtained with 15 different blood tests performed on 510 patients and the age of the patients. The accuracy rate for carrying the disease is 89.6%. An association has been observed between coronavirus and other chronic diseases.

References

[1] Özen, I. A., & Ilhan, I. Opinion Mining in Tourism: A Study on "Cappadocia Home Cooking" Restaurant. In E. Çeltek (Ed.), Handbook of Research on Smart Technology Applications in the Tourism Industry (pp. 43-64). IGI Global, 2020.

[2] Maisha Farzana, Md. Jahid Hossain Any, "Semantic Segmentation of Tumor from 3D Structural MRI using U-Net Autoencoder", Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering, March 2020

[3] Liu W.K., Gan Z., Fleming M., Deep Learning for Regression and Classification. In: Mechanistic Data Science for STEM Education and Applications. Springer, Cham. ,2021

[4] 2019,Available:https://www.bestsoftpro.com/2019/01/a-definition-of-machine-learning-ml.html