Artificial Intelligence and Social Computing, Vol. 28, 2022, 77–84
https://doi.org/10.54941/ahfe1001449

Hepatitis Predictive Analysis Model Through Deep Learning Using Neural Networks Based on Patient History

Remigio Hurtado, Byron Vásquez, Jorge Pizarro, and Willan Mendieta

Universidad Politécnica Salesiana Cuenca, EC 010101, Ecuador

ABSTRACT

First of all, one of the applications of artificial intelligence is the prediction of diseases, including hepatitis. Hepatitis has been a recurring disease over the years as it seriously affects the population, increasing by 125,000 deaths per year. This process of inflammation and damage to the organ affects its performance, as well as the functioning of the other organs in the body. In this work, an analysis of variables and their influence on the objective variable is made, in addition, results are presented from a predictive model. We propose a predictive analysis model that incorporates artificial neural networks and we have compared this prediction method with other classification-oriented models such as support vector machines (SVM) and genetic algorithms. We have conducted our method as a classification problem. This method requires a prior process of data processing and exploratory analysis to identify the variables or factors that directly influence this type of disease. In this way, we will be able to identify the variables that intervene in the development of this disease and that affect the liver or the correct functioning of this organ, presenting discomfort to the human body, as well as complications such as liver failure or liver cancer. Our model is structured in the following steps: first, data extraction is performed, which was collected from the machine learning repository of the University of California at Irvine (UCI). Then these data go through a variable transformation process. Subsequently, it is processed with learning and optimization through a neural network. The optimization (fine-tuning) is performed in three phases: complication hyperparameter optimization, neural network layer density optimization, and finally dropout regularization optimization. Finally, the visualization and analysis of results is carried out. We have used a data set of patient medical records, among the variables are: age, sex, gender, hemoglobin, etc. We have found factors related either indirectly or directly to the disease. The results of the model are presented according to the quality measures: Recall, Precision and MAE. We can say that this research leaves the doors open to new challenges such as new implementations within the field of medicine, not only focused on the liver, but also being able to extend the development environment to other applications and organs of the human body in order to avoid risks possible, or future complications. It should be noted that the future of applications with the use of artificial neural networks is constantly evolving, the application of improved models such as the use of random forests, assembly algorithms show a great capacity for application both in biomedical engineering and in focused areas to the analysis of different types of medical images.

Keywords: Deep learning, Neural networks, Optimization, Fine tuning, Analysis model, Disease prediction, Hepatitis

INTRODUCTION

We will now start with a detailed analysis of hepatitis and its main problem and concern that it has caused within the field of medicine. When the human being appeared, the hepatitis virus was ready to infect human bodies because according to data from science.org, the hepatitis virus is at least 19 million years old, since 2019 there have been 1.1 million deaths from hepatitis worldwide and is increasing.

According to the global study of 'Global Burden of Disease' presented by The Liver Meeting Digital Experience in 2020, few countries are on track to meet the reduction of deaths from viral hepatitis, which is why the elimination of hepatitis is sought by 2030, for which the purpose of this research is to determine values related to the disease, as well as a predictive method to know if there will be problems in the future.

The analysis of these diseases is essential to monitor any relationship with the disease, ASSCAT proposes that any predictive method must be greater than 65% to be considered a safe and reliable model, but since very few countries collect data on this disease, we have compiled a series of data provided by the UCI Machine Learning Repository knowing the fundamentals that have been the basis of the work related to this disease.

Fundamentals

Hepatitis is frequently associated with contact with food or water contaminated by the feces of an infected person, so you can get hepatitis from eating undercooked pork, venison, or shellfish, so the only way to detect these diseases is it is through the symptoms that occur, or in case the patient suffers from a hepatic clinical history. Neural networks have positioned themselves as one of the most fundamental tools to optimize and automate very complicated processes, for which a predictive model for this disease based on neural networks has been developed.

Work Related to Hepatitis

Neural networks can have different applications in the health area. Over the years, many data scientists have dedicated their research to the detection of anti-HCV antibodies, which searches for antibodies against this disease. through the presence of nucleic acids, which amplifies our genetic material, achieving detectable limits.

There are several works carried out to counteract this disease, these focus on supervised learning, decision trees, vector support machines, computational models, neural networks, AI's, etc. Through these solutions or tools, monitoring can be given, as well as detection and diagnosis applied to the patient, the recommendation by the Center for Disease Control of the United States of America (CDC) is to carry out a constant control of people who are at risk of contagion or have genetic factors to be considered risk groups. Among the tools to carry out these tasks are artificial neural networks (ANN), which have the ability to perform classification processes, as well as prediction, these can be applied generating an efficient and effective method.

Table 1. Parameters used in the model.

Variable		Description
epochs		Number of times the forwarding and backpropagation algorithms will be executed (cycles).
batch_size		Amount of data an epoch contains.
optimizer	RMSprop	Optimization algorithm that: Maintain a (discounted) moving average of the square of the gradients. Divide the gradient by the root of this average.
	SGD	Gradient Descent Optimizer
	Adam	Stochastic gradient descent method that is based on the adaptive estimation of the first and second order moments.
activation	relu	Rectified linear drive activation function.
	linear	Linear activation function (pass-through).
	sigmoid	Sigmoid activation function Sigmoid(x) = 1 / (1 + exp(-x)).
k		Number of neighbors
Generations		Number of generations
Father		Number of parents

PROPOSED METHOD

In this section we will find the proposed method in detail, as well as a table of occupied parameters, as well as the specification of the algorithm that made the prediction method possible.

The detailed process and dataset are available on GitHub. URL: https://github.com/VasquezB1/NeuronalNetworks.

Table of Parameters and Neural Networks

In this section we will start by defining in a table some of the parameters and elements present in the process and development of an optimized neural network, followed by a graphical representation of the process of the proposed method.

Table 1 shows a minimum definition of the parameters present in the optimization process of our neural network.

General Diagram of the Model

The parameters used in the neural networks are presented, as well as a detailed process of the steps that made the predictive model possible.

1. Data extraction is to select a dataset where we will find variables, the contribution of the data is crucial and important for the study.

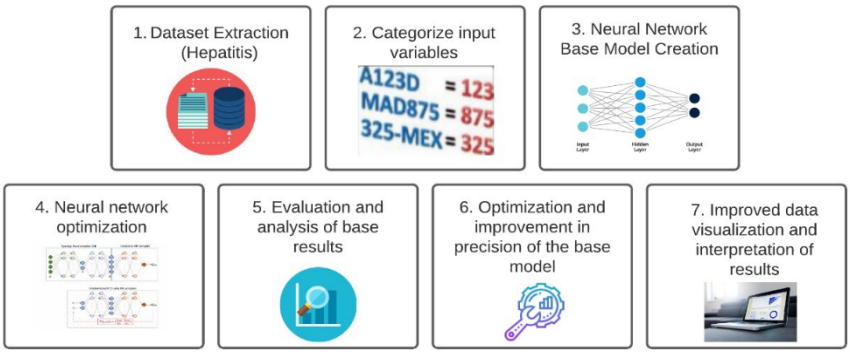

Figure 1: Overview diagram.

2. As a second point we have the change of categorical variables and nominal variables to numerical variables, in order to obtain the same representation, but only in numbers, so that the neural network has a better learning.
3. As a third point, we proceed to the creation of the base neural network, which serves as an example of how much the model can improve.
4. As a fourth point, we start the time function and the execution of the optimization of the neural network to have as a result a better accuracy and a smaller margin of error.
5. As a fifth step, we have the compilation of the program, resulting in an optimized neural network with an acceptable accuracy and margin of error.
6. As a sixth, we have the search for the density of neuron layers, with the aim of optimizing resources and, if necessary, further improving accuracy.
7. We have the visualization of the results obtained as the accuracy and the corresponding margins of error, after due processing and optimization of the neural network.

DESIGN OF EXPERIMENTS

In this section we will find about the descriptions of the data set. Our dataset contains 155 instances and 19 attributes, we will find variables which are grouped by sex, age, antivirals, liver size, varices that the patient has, in addition to the patient's prothrombin time, which directly influences the prognosis. The data set is located within the 'UCI' Automatic Learning Repository, which specifies that through the bilirubin produced by the organism, it can classify said problem and develop a model that helps the rapid response of a positive diagnosis.

A transformation of variables was carried out which, applying a normalization of data, can be eligible to make a prediction model with them, so we will obtain a 0 or 1 in the output, thus indicating the output that it produces.

After this we will create the base model which we will pass as parameters to the neural network, assigning it an amount of nFolds=5 where kfolds is

Table 2. Description of data.

Dataset	Number variables
Number of variables	**19**
Number of records	**155**

Table 3. Variables and attributes used.

Parameters	Results
Batch-size	1, 5, 10, 15, 20, 25, 50, 60, 70, 80, 90,100
Epochs	2, 4, 6, 8, 10, 25, 30, 40, 50
Best optimizer	RMSprop, SGD, Adam

the validation configuration of the K-Fold Cross-Validation, once the base model has been trained, we will assign an evaluation of the results through cross_val_score, where we will obtain a precision of the base model of 0.81%. Now through the optimizer for the model we can improve these results through the training and assignment of neurons to the system, see table 3.

A trained, efficient and more precise model is generated, generating a more favorable result and greater than the 65% required. Each data will pass through a neuron, thus generating a more precise training in addition to looking for its best optimizer and its

best value, avoiding more delay times, being able to visualize a better result. After creating an optimized and trained model in the same way we can evaluate them with cross_val_score, but now passing the new parameters with higher density of neurons which will be activated through the best optimizer.

We will obtain an accuracy of 0.88% with a mean absolute error (MAE) of 0.22%, generating a lower mean error and better processing, we will obtain an improvement of 7.49% in our model after the creation of the neural network trained.

RESULTS AND DISCUSSION

In this section we will find the results generated by the model as well as a comparison between different models such as support vector machines (SVM) and evolutionary computing which, in addition to providing new data, will show us a better option when predicting hepatitis.

Comparison of SVM Models vs Neural Networks

In this section we will compare our neural network model with others applied to the same topic with the prediction of hepatitis on the one hand the SVM, and the evolutionary computation. We will be able to observe that the SVM are applied to classification and regression problems for which they will be perfect to classify our data and to be able to predict directly and effectively, for which we will obtain different values with a different process, but with the same purpose. In the following table we can see the results generated by a support vector machine compared to those of a neural network applied in search of the prediction of hepatitis.

Table 4. SVM vs Neural networks comparison.

Parameters	SVM	Neural network
MAE	0.32	0.22
RMSE	0.56	0.37
ACCURACY	0.67	0.88

Table 5. Results obtained with different tests.

Evidence	RMSE	Accuracy
N° neighbors=5 Generations=50 Parents=3	0.69	0.51
N° of neighbors =10 Generations =500 Parents=3	0.78	0.38
N° of neighbors =5 Generations =100 Parents=6	0.80	0.35

Table 6. Comparative results.

Parameters	SVM	Genetic algorithm	Neural networks
MAE	0.32	0.48	0.22
RMSE	0.56	0.69	0.37
Accuracy	0.67	0.51	0.88

Compared with other models, this one does not have the desired result, because as we explained at the beginning, ASSCAT assigns an average value greater than 65% accuracy so that a model in medicine is valid or can be studied, otherwise the model can reach be detrimental and harmful to both the patient and the doctor.

Comparison of Genetic Algorithm vs Neural Network Models

In this section we will detail the comparison of a genetic algorithm against neural networks with the same purpose of finding a predictive model of hepatitis. We have based ourselves on evolutionary genetic algorithms and machine learning applied to other areas such as marketing and recommendation systems.

After the results obtained, we detect that many processes are not necessary and instead of helping the system, they hinder it and make bigger mistakes. The best result obtained is the one with 50 reproductions, giving us an accuracy value of 0.51, which makes it a very low model compared to neural networks, we have a data dispersion or RMSE of 0.69, which It shows us in the same way that the data is very scattered, generating an unreliable model, as we will see below in the comparative table of results see table 6.

With the results obtained we will be able to create a trend in the values to be recorded, we will realize that some factors already become key for the predictive model, such as age, since it can become a risk factor. In this situation, since if we are ≥ 50 years old, it can affect our results.

CONCLUSION

Through this work we can obtain a prediction about hepatitis, in addition to knowing the different aspects that can affect a positive or negative picture of said disease. We can say that the use of neural networks versus SVM or genetic algorithms is a better solution, in this case we must take into account the training given to each one, they also give us values that can be modified or improved, given that these are in constant development and evolution, these new solutions can be applied to future problems not only related to hepatitis but also as a new model of applications in different fields. For subsequent applications, it is proposed to take the following work as a basis for possible new applications, whether in the field of health or general applications, to carry out an improvement, we recommend the supervision of doctors and their possible diagnoses, thus being key model, for future and present diseases.

ACKNOWLEDGMENT

The authors wish to we would like to thank the Universidad Politecnica Salesiana for providing us with a development and learning environment based on respect and trust, our parents for helping us achieve quality higher education, in addition to providing us with constant strength during this process, as well as the people who supported and made it possible for this work to be carried out successfully.

REFERENCES

Bojorque Chasi, R. X., & Hurtado Ortiz, R. I. (2017). Técnicas híbridas en sistemas de recomendación para optimizar el modelo non negative matrix factorization (Doctoral dissertation, ETSI_Sistemas_Infor).

Barreto, W. and Picón, R. (2020). Estudio experimental y simulación del comportamiento inelástico de paneles compuestos usando redes neuronales artificiales. Informes de la Construcción, 72(558), p.343.

Anon, (n.d.). Diagnóstico de la hepatitis C: Cómo saber si se padece la enfermedad. Pruebas de hepatitis C. Cribado. | ASSCAT. [online] Available at: https://asscat-hepatitis.org/hepatitis-viricas/hepatitis-c/informacion-basica-sobre-la-hepatitis-c/diagnostico-de-la-hepatitis-c-como-saber-si-se-padece-la-enfermedad/

Hurtado, R. (2017). Remigio Hurtado Ortiz Recommender systems clustering using Bayesian non negative matrix factorization. Avalible at: https://scholar.google.com/citations?user=97tZu-YAAAAJ&hl=es&oi=sra

Exsilio Solutions (2016). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog. [online] Exsilio Blog. Available at: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/.

Portal de Salud de la Junta de Castilla y León. (n.d.). Día Mundial contra la Hepatitis. [online] Available at: https://www.saludcastillayleon.es/AulaPacientes/es/dias-mundiales-relacionados-salud/dia-mundial-hepatitis-15f0eb

Descubierto el origen del virus de la hepatitis B. (2010). El País. [online] 30 Sep. Available at: https://elpais.com/sociedad/2010/09/30/actualidad/1285797602_850215.html

Exsilio Solutions (2016). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog. [online] Exsilio Blog. Available at: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/