Konuralp Journal of Mathematics, 9 (2) (2021) 324-331

Konuralp Journal of Mathematics

Research Paper

https://dergipark.org.tr/en/pub/konuralpjournalmath

e-ISSN: 2147-625X

Comparative Analysis of Neural Networks in the Diagnosis of Emerging Diseases based on COVID-19

Murat Kirişci[1*], Ibrahim Demir[2] and Necip Şimşek[3]

[1]*Department of Biostatistics and Medical Informatics, Cerrahpasa Medicine Faculty, Istanbul University-Cerrahpasa, Istanbul, Turkey*
[2]*Department of Statistics, Yildiz Technical University, Istanbul, Turkey*
[3]*Department of Mathematics, Istanbul Commerce University, Istanbul, Turkey*
[*]*Corresponding author*

Abstract

Dermatological diseases are frequently encountered in children and adults for various reasons. There are many factors that cause the onset of these diseases and different symptoms are generally seen in each age group. Artificial Neural Networks can provide expert level accuracy in the diagnosis of dermatological findings of patients with COVID-19 disease. Therefore, the use of neural network classification methods can give the best estimation method in dermatology. In this study, the prediction of cutaneous diseases caused by COVID-19 was analyzed by Scaled Conjugate Gradient, Levenberg Marquardt, Bayesian Regularization neural networks. At some points, Bayesian Regularization and Levenberg Marquardt were almost equally effective, but Bayesian Regularization performed better than Levenberg Marquard and called Conjugate Gradient in performance. It is seen that neural network model predictions achieve the highest accuracy. For this reason, Artificial Neural Networks are able to classify these diseases as accurately as human experts in an experimental setting.

Keywords: *Bayesian Regularization Neural Network; COVID-19; dermatological findings; Levenberg-–Marquardt Neural Network; Scaled Conjugate Gradient Neural Network*

2010 Mathematics Subject Classification: *Primary 03E75; Secondary 03E72, 68T37, 94D05.*

1. Introduction

Artificial Neural Networks are one of the artificial intelligence technologies that simulate the working structure of the human brain, analyze new data, and create new information with different learning algorithms.

Artificial Neural Networks are inspired by biological neurons (nerve cells) and have emerged as a result of artificial simulation studies of the brain's work system. The Artificial Neural Network has started to be used frequently in medical diagnostics recently and it will be seen more in all biomedical systems in the near future. The Artificial Neural Network system is composed of direct, complex, and non-linear models associated with the independent factors of the system and the outputs, which are the predicted factors. The Artificial Neural Network is strong in computing and information processing. Artificial Neural Network derives this power from its parallel structure, ability to learn, and generalize. All these features show that an Artificial Neural Network has the ability to solve complex problems. The neuron, the basic processing element of an Artificial Neural Network, is not linear.

Artificial Neural Networks have two learning types as unsupervised and supervised. The Neural Network is trained according to patterns related to input and output, in supervised learning. During this process, the Neural Networks calculates the weights differently until appropriate results are produced and stop the process when it achieves the most appropriate output. In unsupervised learning, there are only data and no information is given about them.

Generally, feed forward neural networks are trained employing Back Propagation. This operation is the process of learning the relationships between input and output values. The next to work flow in the Back Propagation training algorithm is as follows [1], [2]:

i. Input values propagate towards a hidden layer.
ii. Propagates sensitivities back to minimize the error.
iii. It updates the weights at the end of the process.

Email addresses: mkirisci@hotmail.com (Murat Kirisci), idemir@gmail.com (Ibrahim Demir), necipsimsek@hotmail.com (Necip Simsek)

As a new and highly accurate tool that can assist physicians in analyzing, modeling, and facilitate understanding of complex clinical data in various medical diagnosis practices, Artificial Neural Networks have been used frequently. Most of the Artificial Neural Network applications in medicine are related to classification. That is, the characteristics measured are to include the patient in a smaller class, making it easier to do work [3]. Artificial Neural Networks are useful in simplifying the prognosis, diagnosis, and treatment of many diseases [4]-[17].

In this study, Cutaneous manifestations of diagnosed patients as COVID-19 were investigated with Scaled Conjugate Gradient, Levenberg-–Marquardt, Bayesian Regularization Neural Networks. In this investigation, the prediction capabilities of Artificial Neural Networks were compared. The article is designed as follows: In the second section, applied Artificial Neural Network models are introduced and information about the data used is given. In the third section, the results obtained and the discussions of these results are given. The last section deals with the conclusions.

2. Methods

Type of Study: In this work, dermatological findings of the inpatient and intensive care patients in the hospital of Cerrahpaşa Medical Faculty in the COVID-19 pandemic were investigated.

Selection samples: The database used in the study was created in Cerrahpaşa Medical Faculty. Database records were composed with the knowledge of dermatological findings of inpatient and intensive care patients.

Types of Study Variables: The database consists of eight independent variables and one dependent variable(dermatological findings). Information of 210 patients was taken from the database. The independent variables are in-patient and intensive care patients, gender, age, comorbidity, SARS/COVID/PCR(positive, negative), computed tomography(CT)(positive, negative), time of dermatological findings(before COVID infection, after COVID infection). Variables and their codes are defined in Table 1.

Variables	Definitions(coding)
in hospital	inpatient (1) intensive care (2)
age	18–24 (1) 25–34 (2) 35–44 (3) 45–54 (4) 55–64 (5) 65–74 (6) 75–84 85+ (8)
comorbidity	none (0) one (1) two+more (3)
SARS, COV2, PCR	negative (1) positive (2)
computed tomography	negative (1) positive (2)
before / after COVID infection	none (1) before (2) after (3)
gender	male (1) female (2)

Table 1: Definitions and codes of variables

Methods
In this presented work, Neural Network algorithms were used: Scaled Conjugate Gradient, Levenberg-–Marquardt, Bayesian Regularization Neural Networks.

Artificial Neural Networks propose a different calculation method than known calculation methods. It is possible to see successful applications of this calculation method that adapts to the environment they are in, adaptive, can work with incomplete information, make decisions under uncertainties, and is tolerant of errors, in almost every area of life. In particular, Artificial Neural Networks can be considered among the most powerful techniques in classification, pattern recognition, signal processing, data compression, and optimization studies. It is possible to come across successful examples of Artificial Neural Networks in many areas such as data mining, optical character handling, optimum route determination, fingerprint recognition, material analysis, job scheduling, and quality control, medical analysis.

The idea of artificial neurons was first introduced in 1943 [18]. Since the emergence of this idea, information technologies have progressed rapidly and developed innovative technologies. One of the most important of these innovative technologies is neural networks. In the early 80s, the opinion that a machine could think as a human emerged, and in the early 90s, Artificial Neural Network technologies improved rapidly [18, 19, 20].

Levenberg Marquardt-Back Propagation algorithm

Levenberg Marquardt-Back Propagation algorithm is conceived to application the second-order training method and does not need to use the Hessian matrix. The performance function can be expressed as the sum of squares. Thus the Hessian matrix comes near and the gradient can be calculated:

$$H = M^T M$$
$$g = M^T h.$$

M is a Jacobian matrix in which entries of this matrix are first-order derivatives of network errors according to weights and biases. In this case, the h becomes a vector of network errors. With the standard Back Propagation method, the Jacobian matrix is more easily computable than the Hessian matrix. The Levenberg Marquardt-Back Propagation algorithm employs the Hessian matrix approach in the

$$x_{k+1} = x_k - [M^T M + \rho I]^{-1} M^T h \tag{2.1}$$

Newton-like equation. x in this equation indicates connection weights. If ρ is selected as zero, Equation 2.1 is Newton's method, which uses the approximation Hessian matrix. ρ large can also be chosen. In this case, the equation will be gradient descent with a small step-size. In order to obtain the minimum error, Newton's method is faster and more correct. Therefore, objective is to move on Newton's method. After successful steps, ρ is minimized. When ρ increases, it will increase the tentative step performance function. Therefore, every time, the performance function will be decreased with every iteration. Considering the gradient descent technique, it can be said that the Levenberg Marquardt technique is more powerful.

Scaled Conjugate Gradient-Back Propagation algorithm

In a fundamental Back Propagation algorithm, weights regulate in the steepest descent direction, which is the most negative of the gradient. In this direction, the performance function is decreasing the fastest. This does not create generate the most rapidly convergence even if the function decreases the fastest along with the negative of the gradient. Conjugate gradient algorithms preserve the error minimization obtained in all preceding steps. In this case, a search is usually implemented along such a direction producing most rapidly convergence than the steepest descent direction. Conjugate direction is the naming given for this direction. Regulating the step-size at every iteration is existing in almost every Conjugate Gradient algorithm. Determining the step-size is a process performed along the Conjugate Gradient direction and the performance function will be minimized during this process. Starting by searching in the steepest descent direction on the first iteration is the main characteristic of Conjugate Gradient algorithms (Equation 2.2). It is very common to use Conjugate Gradient algorithms in line search (Equations 2.3, 2.4).

$$p_0 = -g_0 \tag{2.2}$$
$$x_{k+1} = x_k + \alpha_k g_k \tag{2.3}$$
$$p_k = -g_k + \beta_k p_{k-1} \tag{2.4}$$

The way in which factor β_k is calculated allows one to distinguish Conjugate Gradient algorithms.

There are also different approaches to estimating step-size except for the line search method. By integrating the model trust region way known from the Levenberg-–Marquardt algorithm with Conjugate Gradient, a new technique can be created. The new technique is called Scaled Conjugate Gradient. As it is described in

$$s_k = \frac{E^{'}(w_k + \sigma_k p_k) - E^{'}(w_k)}{\sigma_k} + \lambda_k p_k, \tag{2.5}$$

in this approach, scaling factors λ_k and σ_k are acquainted with approximate the Hessian matrix and commenced by the user at the starting of the algorithm such that $0 < \lambda_k < 10^{-6}$ and $0 < \sigma_k < 10^{-4}$, where s is Hessian matrix approximation, E is the total error function and $E^{'}$ is the gradient of E.
For Scaled Conjugate Gradient, factor β_k computation and direction of new search are as follows:

$$\beta_k = \frac{(|g_{k+1}|^2 - g_{k+1}^T g_k)}{g_k^T g_k} \tag{2.6}$$
$$p_{k+1} = -g_{k+1} + \beta_k p_k. \tag{2.7}$$

For the success of the algorithm, it is very important to update the design parameters independently at each iteration user. This is a great advantage considering the operation of line search-based algorithms.

Bayesian Regularization-Back Propagation algorithm

Bayesian Regularization is an example of a training algorithm in which weight and bias values are updated. Bayesian Regularization minimizes the composition of weights and squared errors and after that identifies a accurate composition to conceive a network that can be best generalized. Bayesian Regularization presents network weights into the second-order training objective function which is shown as $F(\omega)$ in Equaiton 2.8:

$$F(\omega) = \alpha_\omega + \beta E_D \tag{2.8}$$

In this equation, the sum of network errors, the sum of the squared network weights, and objective function parameters are denoted E_D, E_ω, α, and β, respectively. In Bayesian Regularization, the weights of the network are random variables. In this case, second-order training set and the distribution of net weights is taken as the Gaussian distribution.

Let's use Bayes' theorem while defining α and β factors. For the variables U and V:

$$P(U|V) = \frac{P(V|U)P(U)}{P(V)} \tag{2.9}$$

IN 2.9, the posterior probability of U conditional on V, the prior of V conditional on U and the non-zero prior probability of event V are denoted $P(U|V)$, $P(V|U)$ and $P(V)$, respectively.
Where $P(U|V)$ is the posterior probability of U conditional on V, $P(V|U)$ the prior of V conditional on U, and $P(V)$ the non-zero prior probability of event V, which functions as a normalizing constant. Minimizing the objective function (Equation 2.8) is necessary to find the optimum weight. This means that the posterior probability function

$$P(\alpha,\beta U|W,M) = \frac{P(W|\alpha,\beta,M)P(\alpha,\beta|M)}{P(W|M)} \tag{2.10}$$

is maximized. In 2.10, the factors needed be to optimized, the weight distribution, the particular neural network architecture, the normalization factor, uniform prior density for the regularization parameters and the likelihood function of W for given α,β,M are denoted α and β, W, M, $P(W|M)$, $P(\alpha,\beta|M)$, and $P(W|\alpha,\beta,M)$, respectively.
If we maximize the posterior function $P(\alpha,\beta|W,M)$, it means that the likelihood function $P(W|\alpha,\beta,M)$ is maximized. After doing this operation, optimal values for α and β to a certain weight space are obtained. Subsequently, the algorithm moves into Levenberg-–Marquardt stage where Hessian matrix computations occur and update the weight space in order to minimize the objective function. Thereafter, if the convergence is not met, the algorithm anticipates new values for α and β and the whole transaction recaps itself until convergence is achieved.

Analyses

The network model was created using MATLAB(2015a) neural network toolbox. The training algorithms targeted in this work and the MATLAB functions used for these algorithms are explained in Table 2.

Table 2: Evaluated training Algorithms

Training Function	Description
trainlm	LMLevenberg-–Marquardt
trainbr	Bayesian Regularization
trainscg	Scaled Conjugate Gradient

A Neural Network consisting of layers (1 input, 1 hidden, 1 output) is presented 2.1. The number of neurons in the hidden layer of each selected Training Algorithm is collimated to achieve the best performance. The intent of the training and validation phases is to compose an optimal weight space in order to build the mapping of the taken out noise constituents from input and target datasets. A Neural Network is created and then pre-normalized datasets are distributed across the network. The dataset is disunited into the subgroups as training, validation, Testing. In the training stage, Training Algorithms try to accurately the randomly distributed initial weight space until the performance objective of the validation stage is recorded or no further correction is required after several consecutive iterations. The validation set is established to prevent the overfitting on the training data, as an Artificial Neural Network without a validation set is probable to be overfitted on the training data.If overfitting happens, then the network mislays its ability to discover a fundamental connection between training and testing sets, rather than focalize on the training set performance, which brings the Testing set performance significantly lower. Levenberg-–Marquardt, Scaled Conjugate Gradient and Bayesian Regularization-Back Propagation Training Algorithms with configuration parameters as given in Table 3 are used to train the network.

Table 3: Definitions and codes of variables

Parameters	Scaled Conjugate Gradient	Levenberg-–Marquardt	Bayesian Regularization
Maximum number of people to train	1000	1000	1000
Performance goal	0	0	0
Maximum validation failures	2	2	2
Initial ρ	N/A	0.001	0.005
ρ decrease factor	N/A	0.1	0.1
ρ increase factor	N/A	10	10
Maximum ρ	N/A	1e10	1e10
σ_0	5e-5	N/A	N/A
λ_0	5e-7	N/A	N/A

In these models established, the number of hidden neurons was taken as 10 for each neural network.(Figure 2.2). The data are divided as follows: training 70%, validation 15%, testing 15%.

3. Results and Discussion

This work, it is aimed to create models that can predict cutaneous diseases caused by COVID-19 with different Back Propagation Training Algorithms and to select the most suitable of these models. The data was created as a dataset containing the dermatological findings of patients in the COVID-19 service. This dataset contains 210 patients. 142 samples were used in training the network. The dermatological disease was defined as ”exist” and ”none” in each patient. Dermatological findings take binary values: 1 indicates that it is non-disease and 2

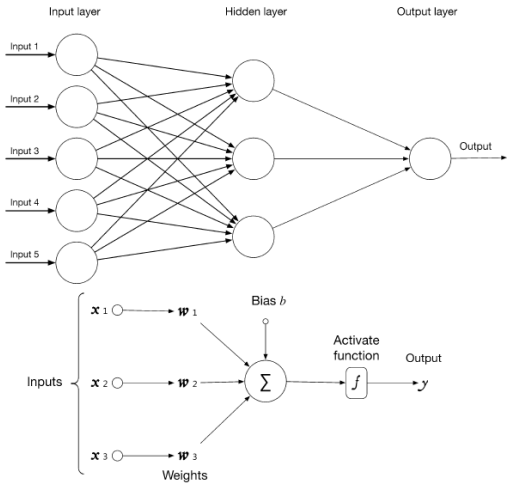

Figure 2.1: Neural Network Structure

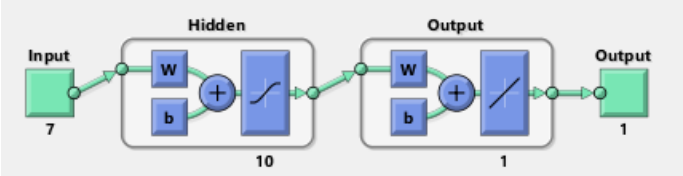

Figure 2.2: Model

indicates that there are diseases.

The status of the variables in the data set is given in Table 4. The age range and frequencies in Dataset are shown in Table 5 (Figures 3.1-3.3).

Table 4

Gender	male	123
	female	87
in-patient/intensive care	in-patient	163
	intensive care	47
comorbidity	none	88
	one	48
	two + more	74
SARS/COV2/PCR	negative	88
	positive	122
computed tomography	negative	8
	positive	202
After/Before COVID infectious	after	52

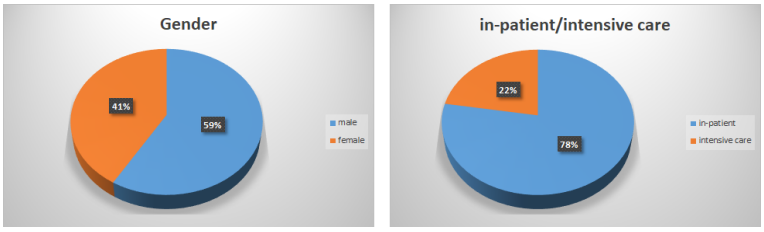

Figure 3.1: Gender and in-patient/intensive care

Table 5: Frequency of Age

Range of age	Frequencies
18–24	5
25–34	18
35–44	26
45–54	34
55–64	50
65–74	43
75–84	27
85+	7
total	210

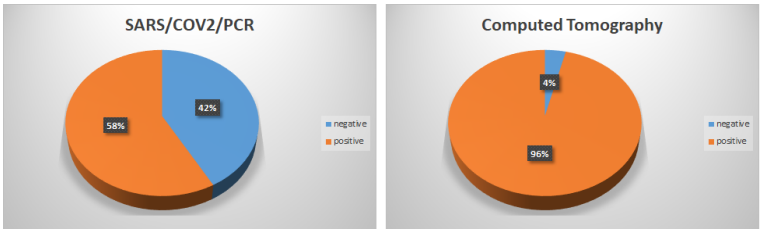

Figure 3.2: SARS/COV2/PCR and Computed Tomography

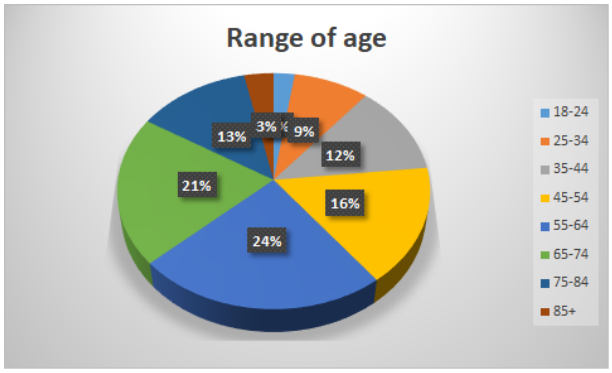

Figure 3.3: Range of age

The plots of performance and regression of these algorithms are indicated in Figures 3.4 and 3.5. In Figure 3.4, the best MSE for training, testing, and validation are displayed. Figure 3.5 indicates the R-squared values for training, Testing, validation.

A trial and error method was made on the network and it was decided how many neurons would be in the hidden layer with this method. Optimal results were achieved in 10 neurons in all three algorithms. R-squared values were found for Scaled Conjugate Gradient, Levenberg-–Marquardt, and Bayesian Regularization algorithms as 0.99596, 1, and 0.99905, respectively. It is understood from the obtained R-squared values that there is a very good correlation between observed and predicted values. That is, it explains that the Artificial Neural Network algorithms applied to the data are a very suitable model for predicting optimal conditions for dermatological findings based on COVID-19.

According to the results obtained when the Artificial Neural Network methods were applied, they were revealed that the network performed very well in learning patterns related to the symptoms of diseases. The network was simulated in situations that it has not encountered before, namely in the test data. The results are great(Figure 3.5). MSE is the average square difference between outputs and targets. A zero value means that there are no errors and a low value means the best result. This kind of net can comply with multi-dimensional mapping problems arbitrarily well, given coherent data, and sufficient neurons in its hidden layer(Figure 2.2).

The outputs were compared and it was seen that Bayesian Regularization was better than Levenberg-–Marquardt and Scaled Conjugate Gradient from the results obtained (Figure 3.4). Thus, $BayesianRegularization > Levenberg--Marquardt > ScaledConjugateGradient$. The fact that $R^2 = 1$ means that an Artificial Neural Network definitely gives better results. The fact that the Bayesian Regularization has performed well on the dataset explains that by applying regularization, the over-fitting problem in Bayesian Regularization is eliminated. The reason for this is due to the probabilistic nature of Bayesian Regularization.

The corresponding performance plot and regression plot of Bayesian Regularization, Levenberg-–Marquardt, and Scaled Conjugate Gradient algorithms generated by MATLAB were examined and demonstrated in Figures 3.4, 3.5. Figures 3.4 (a), (b), and (c) described the best MSE for training, Testing, and validation for the Scaled Conjugate Gradient, Levenberg-–Marquardt, Bayesian Regularization networks respectively. Figure 3.4 (c) showed that the best training performance was Bayesian Regularization with 95 epochs.
R-square values for Scaled Conjugate Gradient, Levenberg-–Marquardt, and Bayesian Regularization, training, Testing, validation, respec-

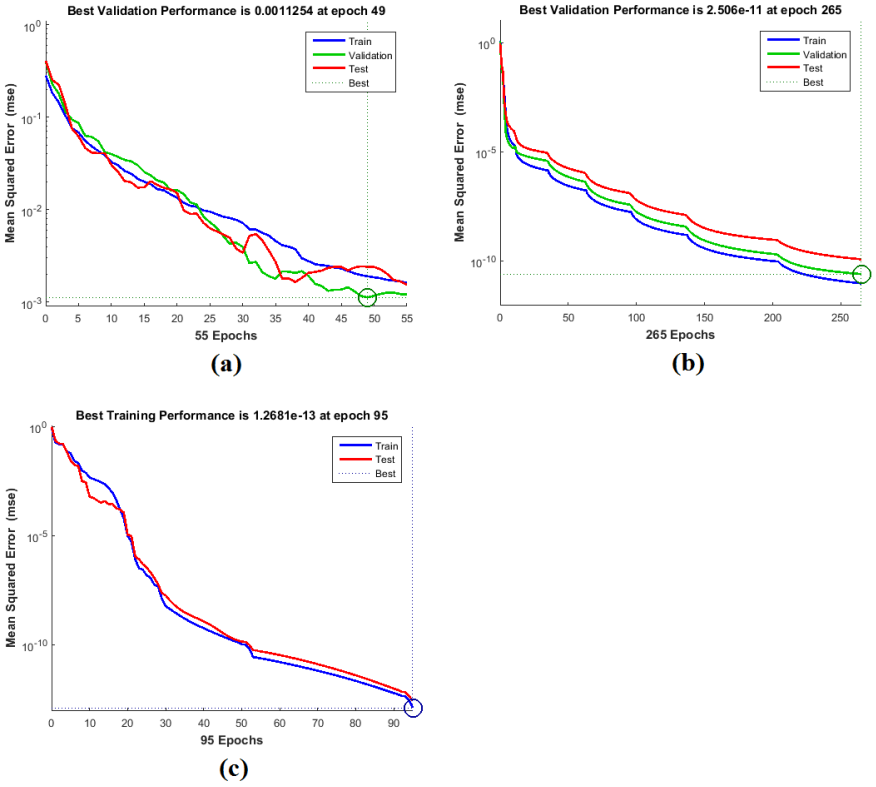

Figure 3.4: Graph of the best performance: (a) Scaled Conjugate Gradient (b) Levenberg-–Marquardt (c) Bayesian Regularization

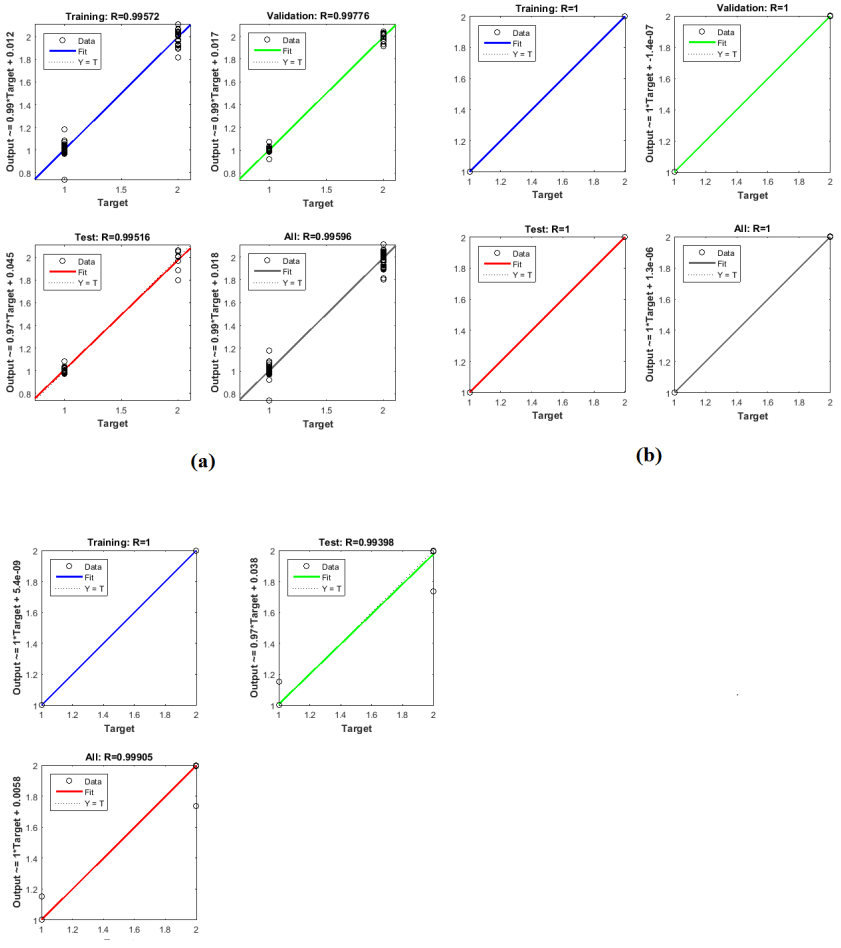

Figure 3.5: R values: (a) Scaled Conjugate Gradient (b) Levenberg-–Marquardt (c) Bayesian Regularization

tively, are shown in Figure 3.5 (a), (b), (c). For this reason, it can be said that Levenberg-–Marquardt is the most appropriate model for predicting optimal situations in the detection of dermatological diseases caused by COVID-19.

4. Conclusion

The intent of this paper is to focus on dermatological diseases in patients with COVID-19 disease as its research theme. The performance of an Artificial Neural Network in obtaining dermatological findings in COVID-19 patients was evaluated. These results suggest that Artificial Neural Network approaches can more accurately identify COVID-19 patients incorporating dermatological diseases. That is, these outcomes demonstrate that the proposed method has the capability to procreate efficient rules with the highest accuracy for the detection of dermatological findings of COVID-19 patients. Artificial Neural Networks can classify dermatological diseases at a level similar to dermatologists. This proposed method predicts that physicians can help differentiate dermatological findings in diseases such as COVID.

In conclusion, it was found that a hidden layer acquired the best classification performance provided that it was placed with neurons on the hidden layers. In summary, the obtained results proved that the Artificial Neural Network model is a good choice for performing classification.

Article's information

Some of the results of this study presented in 10*th* International Eurasian Conference on Mathematical Sciences and Applications (IECMSA-2021).

Funding Information

This study was not funded by any institution.

Conflict Interest

The authors declare that no competing interests exist.

References

[1] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw., 5 (1994), 989–993.
[2] Murat Kayri, Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data, Math. Comput. Appl. 21, (2016), 20, DOI:10.3390/mca21020020
[3] R. Dybowski, V. Gant, Clinical Applications of Artificial Neural Networks, Cambridge University Press, 2007.
[4] V. Arbabi, B. Pouran, G. Campoli, H. Weinans, A.A. Zadpoor, Determination of the mechanical and physical properties of cartilage by coupling poroelastic-based finite element models of indentation with artificial neural networks, J Biomech. 49, (2016), ,631–637.
[5] J. Dolz, L. Massoptier, M. Vermandel, Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: a survey, IRBM, 36(4), (2015), 200–212.
[6] O. Er, N. Yumusak, F. Temurtas, Chest disease diagnosis using artificial neural networks, Expert Systems with Applications, 37, (2010), 7648–7655.
[7] O. Er, F. Temurtas, A.C. Tanrikulu, Tuberculosis Disease Diagnosis Using Artificial Neural Networks, J Med Syst, 34, (2010), 299–302, DOI: 10.1007/s10916-008-9241-x
[8] J.M. Haglin, G. Jimenez, A.E.M. Eltorai, Artificial neural networks in medicine, Health and Technology, 9, (2019), 1–6, DOI: 10.1007/s12553-018-0244-4
[9] M. Kirisci, H. Yilmaz, M.U. Saka, An ANFIS perspective for the diagnosis of type II diabetes, Annals of Fuzzy Mathematics and Informatics, 17, (2019), 101–113, DOI: 10.30948/afmi.2019.17.2.101
[10] M. Kirisci,Comparison of artificial neural network and logistic regression model for factors affecting birth weight, SN Appl. Sci., 1, (2019), 378, DOI: 10.1007/s42452-019-0391-x
[11] M. Kirisci, H. Yilmaz, Artificial Neural Networks for Analysis of Factors Affecting Birth Weight, American Journal of Information Science and Computer Engineering, 5, (2019), 17–24.
[12] J. Kojuri, R. Boostani, P. Dehghani, F. Nowroozipour, N. Saki, Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram, Journal of Cardiovascular Disease Research, 6(2), (2015), 51–59.
[13] C.C. Lin, Y.K. Ou, S.H. Chen, Y.C. Liu, J. Lin, Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture, Injury. 41(9), (2010), 869–873.
[14] I. M. Nasser, S.S. Abu-Naser, Predicting Tumor Category Using Artificial Neural Networks, International Journal of Academic Health and Medical Research, 3, (2019), 1–7.
[15] T. Nowikiewicz, P. Wnuk, B. Małkowski, A. Kurylcio, J. Kowalewski, W. Zegarski, Application of artificial neural networks for predicting presence of non-sentinel lymph node metastases in breast cancer patients with positive sentinel lymph node biopsies, Archives of Medical Science, 13, (2017), 1399-–1407.
[16] M. Shioji, T. Yamamoto, T. Ibata, T. Tsuda, K. Adachi, N. Yoshimura, Artificial neural networks to predict future bone mineral density and bone loss rate in Japanese postmenopausal women, BMC Research Notes, 10, (2017), 590.
[17] H. Yu, D.C. Samuels, Y.Y. Zhao, Y. Guo, Architectures and accuracy of artificial neural network for disease classification from omics data, BMC Genomics, 20, (2019), 167, DOI:10.1186/s12864-019-5546-z
[18] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bulletin of Math. Biophysics, 5, (1943), 115–133.
[19] J. j. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Sci., 79, (1982), 2554–2558.
[20] T. Kohonen, Self-organised formation of topologically correct feature maps, Biological Cybernetics, 43, (1982), 59–69.