The Journal of Supercomputing (2024) 80:4533–4553
https://doi.org/10.1007/s11227-023-05635-z

Optimization of artificial neural network structure and hyperparameters in hybrid model by genetic algorithm: iOS–android application for breast cancer diagnosis/prediction

Mehmet Akif Bülbül[1]

Accepted: 28 August 2023 / Published online: 14 September 2023

Abstract
Breast cancer is a common disease that can result in death among women. Cancer research is important because early detection of cancer facilitates clinical practice for patients. The aim of the study is to ensure that breast cancer can be diagnosed in a short time and easily. For this purpose, a dataset containing 116 samples, 9 features and 2 target variables (Breast Cancer Coimbra) from the UCI library was used during the training and testing phases. A hybrid structure was created with genetic algorithm (GA) and artificial neural network (ANN) to classify the datasets. With the established hybrid model, the feedforward backpropagation artificial neural network model and the hyperparameters in this model structure have been optimized with the genetic algorithm. The performance of the structure constructed with the most successful gene parameters obtained was compared with weighted K-nearest neighbors, decision tree, and linear support vector machine methods. In all machine learning methods used, fivefold cross-validation was applied and the dataset was divided into two groups as 50% training and 50% testing in order to test the models with different data. The hybrid model proposed in the study performed better than other machine learning methods with 100% correct classification rate. Although there are few data in this study, the accuracy is higher than other literature. In addition, an iOS–android-based application has been developed for the diagnosis and prediction of the disease with the findings obtained. Thanks to the developed application, the most important factor in the fight against the disease, time and cost spent for the diagnosis of this disease will be saved. Considering the interest in artificial intelligence techniques in cancer research, this study presents a new diagnostic method and a usable application in terms of patient decision support systems.

Keywords Breast cancer classification · Model and hyperparameter optimization · Genetic algorithm · Artificial neural networks · Decision support systems

Extended author information available on the last page of the article

1 Introduction

One essential factor that can cause death in women is breast cancer [1]. The mortality rate from breast cancer is 48% in nations with low or moderate levels of the Human Development Index. As compared to nations with high or very high human development metrics, this rate is four times greater. The early identification of breast cancer is rendered possible by good screening programs, which benefits the treatment of breast cancer. [2] The main method for diagnosing breast cancer is mammography. But nevertheless, women with thick breast tissue's scanning sensitivity decline for causes such as tissue overlap [3]. Early diagnosis and treatment of breast cancer have been shown to be beneficial in eradicating the disease [4]. Moreover, early identification and treatment are crucial because of the substantial economic and societal costs associated with breast cancer.

Many artificial intelligence techniques have been applied to anticipate and diagnose patients with breast cancer as a result of technological advancements. For the diagnosis and classification of breast cancer, Lin et al. [5] used artificial neural networks (ANNs) and support vector machines (SVMs). In the study, SVM performed a more accurate classification, and in a subsequent step, transfer learning was used to train the AlexNet, ResNet101, and InceptionV3 networks. Successful outcomes for the diagnosis of breast cancer were determined by the research.

Rasool et al. [6] used the logistic regression (LR), K-nearest neighbor (KNN), community classifier (EC), and SVM techniques to try and identify benign and malignant tumors. These models' diagnostic ability was enhanced using data exploratory techniques. SVM polynomial kernel had a success rate of 99.3% in the research using Wisconsin Diagnostic Breast Cancer (WDBC) and Breast Cancer Coimbra Dataset (BCCD).

For the diagnosis of breast cancer, Jayandhi et al. [7] created an effective deep learning architecture (DLA) with SVM. In this experiment, which combines concepts from DLA and SVM, tiny size 3x3 convolution filters and 16-layer state-of-the-art visual geometric group (VGG) architecture are employed, which decreases system complexity. The findings of the investigation demonstrated that the VGG-SVM model successfully classified images from the Mammographic Imaging Analysis Association database with 98.67% accuracy, 99.32% sensitivity, and 98.34% specificity.

So for diagnosis of breast cancer, Gutierrez et al. [8] applied deep learning. In the study, which also addresses the false positive concern in breast cancer diagnosis, Resnet34, Tsallis entropy-based image segmentation, and Gaussian filter-based preprocessing approaches were selected for feature extraction. The chimpanzee optimization algorithm was used in the research to successfully tune the ResNet 34 model's hyperparameters (COA).

In their investigation on the detection of breast cancer, Rani et al. [9] employed decision tree and average sensor algorithms. Several ML techniques were experimentally tested in this work, which used algorithms to develop a hybrid structure. The proposed hybrid technique achieved an accuracy of 98.1% as a result of the analyses and assessments.

For the early identification of breast cancer, Tarawneh et al. [10] suggested a decision tree-based methodology. Ten breast cancer samples from the Kaggle archive were used in the study, and 286 breast cancer samples from the same pool were used in the follow-up. In the study, the first trial's success rate was 100%, while the second trial's success rate was 97.9%.

For the diagnosis of breast cancer, Aslan et al. [11] employed ANN, conventional extreme SVM, KNN, and extreme learning machine (ELM) techniques. In the study, various ML techniques were applied to the BCCD dataset. The experimental experiments led to an accuracy of 83.80% with the ELM approach.

In order to diagnose breast cancer, Fijri et al. [12] used fuzzy core c-means (FKCM), clustering, and sparse learning fuzzy c-means algorithms. The study that employed various ML techniques used the BCDD dataset. They achieved 89.20% accuracy with the FKCM approach in the study in which the experimental studies were presented in a comparative manner.

An expert system for use in the diagnosis of breast cancer was described by Araújo et al. [13]. Based on the ideas of neural networks and fuzzy systems, they proposed a hybrid model in the study. In this study, the fuzzy neural network model was successful in diagnosing breast cancer 81.04% of the time using the BCDD dataset.

An efficient predictive model for use in the detection of breast cancer was developed by Yavuz and Eyupoglu [14] using a generalized regression neural network model with median filtering and a stepwise principal component analysis technique. Using the BCDD dataset, the study had a success rate of 97.73%.

Breast cancer classification was suggested by Alshutbi et al. in 2022 [15]. In the suggested study, the Jaya algorithm was used to optimize the penalty factor parameter and kernel parameter in the SVM method. Using the BCDD dataset, the Jaya-SVM hybrid structure has a success rate of 98.21%.

Most studies in the literature end up choosing their models for use in artificial intelligence techniques through a process of trial and error. The success of the applied model is directly impacted by changes in the number of input parameters utilized for the chosen method or the hyperparameters determined in the installed architectures. As a consequence, the models' hyperparameters should be optimized in accordance with the input parameters, and the hyperparameters of the models utilized in accordance with the various input parameters should be optimized independently. Such nested optimization problems can only be solved by developing hybrid structures utilizing ML models and optimization algorithms [16, 17]. The optimum outcome, or a solution that comes near to the best solution, for the issue is provided by hybrid structures created employing optimization methods [18].

In this research, a hybrid structure was developed for breast cancer diagnosis and prediction, and a model was developed by optimizing all hyperparameters in a feed-forward backpropagation ANN structure. On BCDD, the suggested model was used. The dataset that was not included in the training was tested using several ML techniques and the suggested hybrid model, and the results were published. Also, the model's outcomes using the hybrid structure are shown and examined in relation to previous research. Also, an application built for the iOS and Android platforms has

been created to aid in the diagnosis and forecasting of the condition utilizing the parameters gleaned from experiments conducted.

2 Materials and methods

2.1 Genetic algorithm

Artificial intelligence optimization algorithms are constantly striving to enhance the solutions, albeit they cannot guarantee that the solution they find is the best solution but rather that it is very close to the optimum solution [18]. GA is one of the optimization methods that is often used in the literature and aims to discover the best answers to problems by replicating natural processes. Figure 1 illustrates the genetic algorithm's flowchart.

In Fig. 1, GA produces a population of random solutions. Each solution in the population is a binary sequence. A chromosome is the term of this structure. Each chromosome's fitness is assessed using the fitness function [19]. The chromosomes in the population are subjected to many genetic processes, including crossover, mutation, and selection.

2.2 Artificial neural network

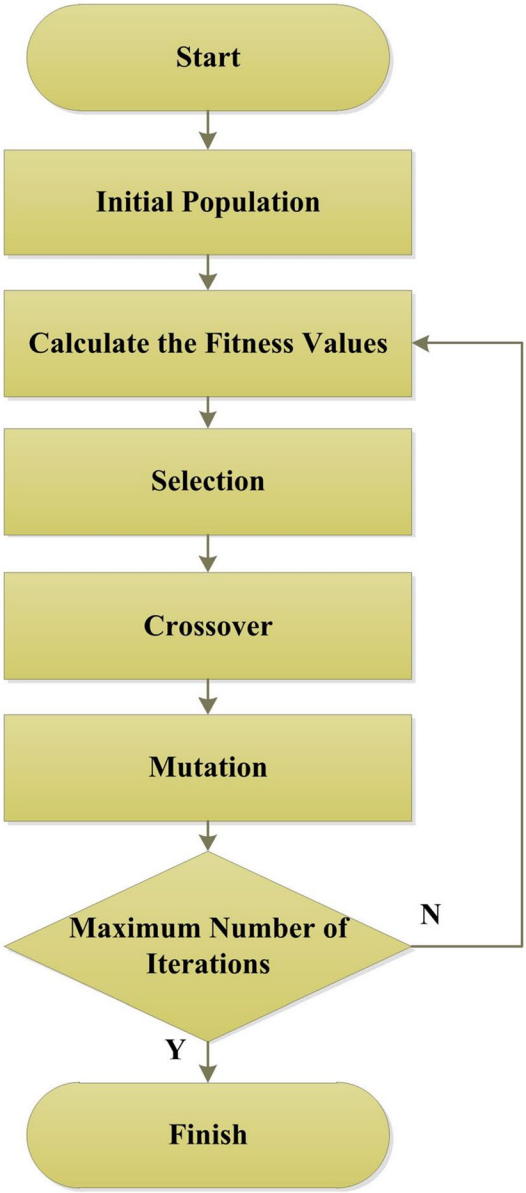
An information ML technique called ANN uses the function of nerve cells as an illustration [17]. Artificial neural network modeling for steam ejector design requires rigorous data processing to produce predictions and models that faithfully mirror real systems. More than one nerve cell connecting to one another creates network architectures. Figure 2 depicts the structure of a single neuron cell.

In Fig. 2, the inputs to the neuron cell are multiplied by their weights before being sent to the addition function. The bias is applied to the result of the addition function. The bias allows the activation function to be shifted because of the various value it accepts. Consequently, the neuron's net input is determined. By processing this net input through the activation function, the neuron cell's output is obtained.

Several algorithms are utilized in the literature to train the ANN network structure [20]. Backpropagation algorithms in feedforward networks have been used successfully in supervised learning algorithms [21]. The illustration in Fig. 3 shows a basic feedforward backpropagation ANN structure.

The outputs of each neuron cell are utilized as the input data for the subsequent neuron cell in the feedforward and backpropagation network structure shown in Fig. 3. The raw input data sent to the network are located in the input layer. The input data are multiplied by the weights chosen by the network when it is sent to the hidden layer. By adding the bias values chosen by the network, these values which are summed up in the sum function in each neuron cell are transferred to the activation function. With the activation function, the neuron's input information is transformed into output information. The same techniques are used to send the output data collected from each neuron cell across the network to the output

Fig. 1 GA flowchart



layer. By comparing the output information produced at the output layer with the desired output information, the error is computed. This 'error' is mirrored in the network in the back propagation network structure by the learning function, allowing the weights to be updated once more. The number of inputs employed by each of the three levels of the network structure shown in Fig. 3, the number of

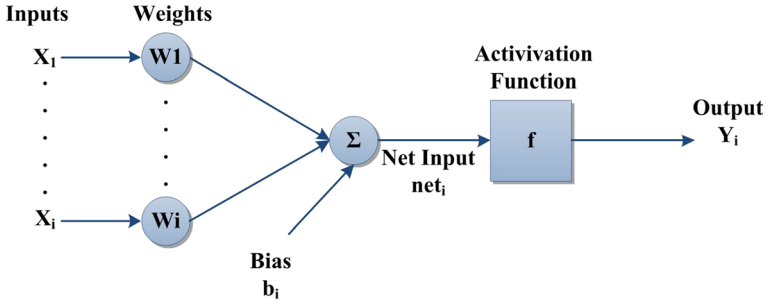


Fig. 2 Neuron cell structure

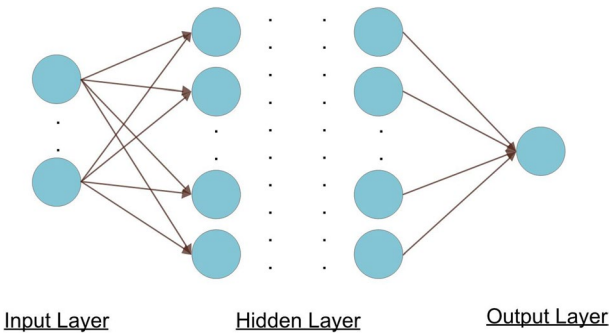


Fig. 3 A feedforward backpropagation ANN structure

hidden layers, the choice of activation function, and the learning function all have an impact on how well the network performs.

2.3 Other machine learning methods used in the study

Table 1 lists additional machine learning techniques that were applied in this research.

Machine learning techniques presented in Table 1 are frequently used in the literature [28–31].

3 Experimental studies and results

3.1 Dataset

Breast cancer was identified in this research using the Breast Cancer Coimbra (BCDD) dataset, which is available at UCI. The BCDD dataset is frequently utilized in the literature [6, 32, 33]. This dataset contains nine defining features, which are listed in Table 2.

Table 1 Other machine learning methods used in the study

Machine learning method	Description
Weighted KNN	The weights of the features are taken into consideration while computing each object's KNN using WKNN. The class to which the categorical data will belong is decided using the distance-weighted KNN rule [22–24]
Decision Trees	Testing, branching, and data categorization procedures are all included in the decision tree. In order to reach the leaves, it moves from the root node. The data for leaves are classified [25, 26]
Support Vector Machines	With regard to regression and classification problems in machine learning, support vector machine (SVM) is a potent approach. Its goal is to identify the hyperplane that splits the space-based data points most effectively. The objective of this plane is to effectively divide data from various classifications [27]

Table 2 Descriptive attributes found the dataset

No	Decisive attributes	Unit of
1	Age	year
2	BMI	kg/m^2
3	Glucose	mg/dL
4	Insulin	μU/mL
5	Homeostasis Model Assessment	HOMA
6	Serum value of Leptin	ng/mL
7	Adiponectin	μg/mL
8	Resistin	ng/mL
9	Chemokine Monocyte Chemoattractant Protein 1 (MCP-1)	(pg/dL)

Between 2009 and 2013, the Physiology Laboratory of the Faculty of Medicine at Coimbra University in Portugal used the data acquired by taking different measures from peripheral venous blood bottles taken in the hospital for all subjects. A pathologist trained in the Department of Pathology examined tissues collected from cancer patients. [34]. The collection contains 116 pretreatment samples from 64 breast cancer patients and 52 healthy women [35]. Healthy women were classed as 1, and unwell women as 2, in the dataset.

In order to reduce the number of samples in the dataset and to test the models more thoroughly using various constructs, the dataset was split into two groups: 50% training data and 50% test data. In order to evaluate the performance of the models to be used in experimental research on the data that are not seen, as objectively and accurately as feasible, fivefold cross-validation was employed throughout the training stages.

3.2 Performance evaluations matrices

In this research, four cross-validation matrices for each classifier were compared. They are precision, recall, F1 score, and accuracy. A confusion matrix is used to calculate these values. Figure 4 illustrates a confusion matrix structure.

The true positive (TP) in the confusion matrix shown in Fig. 4 stands for people who are believed to be healthy. False negative (FN) is the percentage of healthy people who are thought to be ill. False positive (FP) is the number of people who are thought to be healthy but are actually sick. The number of sick people projected to be sick is represented by true negative (TN). Equations 1–4 display the computation of the precision, recall, F1 score, and accuracy values in accordance with the matrix displayed in Fig. 4 [6].

$$\text{Precision}\,(P) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall}\,(R) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1 score} = \frac{2 * P * R}{R + R} \tag{3}$$

$$\text{Accuracy}\,(A) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{4}$$

3.3 Proposed model

In this part of the research, a hybrid model combining GA and feedforward backpropagation ANN is suggested to be utilized for the detection and prediction of breast cancer. Many hyperparameters in the ANN network layout have a direct impact on the network performance. To achieve successful outcomes, it is crucial to optimize these hyperparameters. The success of the network is also strongly impacted by variations in the input parameters that must be provided as input. For various inputs, the same network architecture yields various results. The input data should also be optimized while choosing the network topology with the best

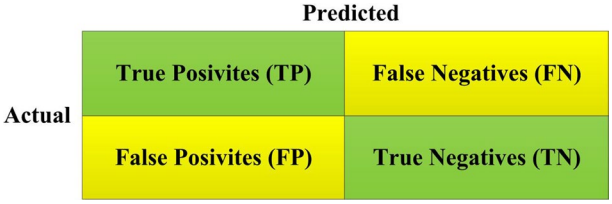


Fig. 4 Confusion matrix structure

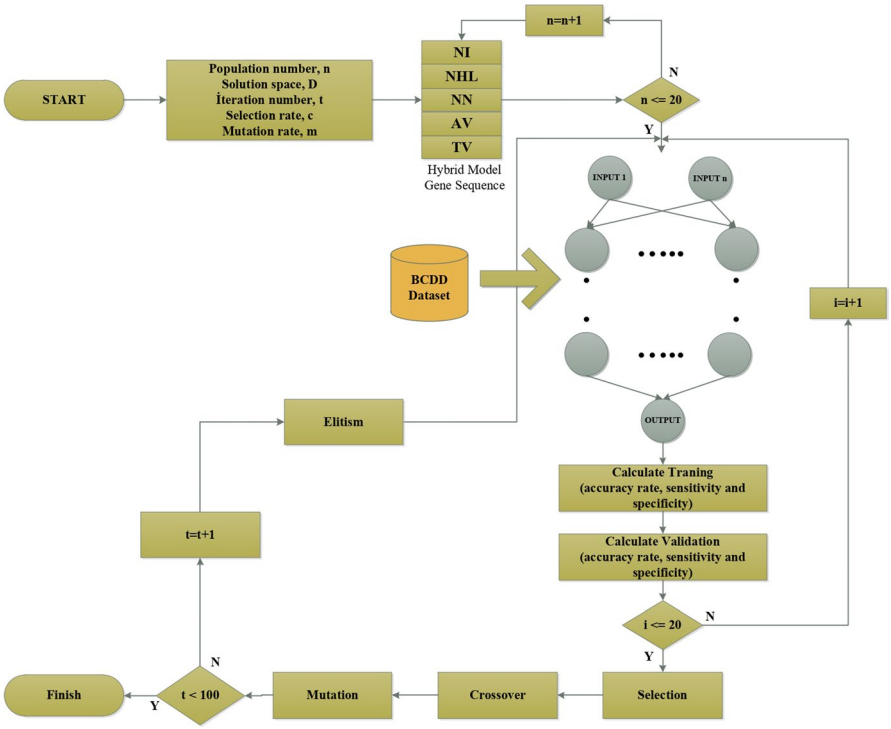


Fig. 5 Hybrid model flowchart

Table 3 GA parameters for hybrid model

Algorithmn parameters	Values
Population number	20
Solution space	5
Selection rate	0.85
Mutation rate	0.03
İteration number	100

hyperparameters for the challenge at hand. Optimization methods can be used to find the ANN network structure with the best or nearly optimal performance [16] Fig. 5 depicts the flowchart of the suggested hybrid structure for diagnosing and predicting breast cancer.

The initial parameters of the model described in Fig. 5 were established in the first stage of the flow diagram, as stated in Table 3. The MATLAB platform was used to create the model due to its advantages for coding and structural flexibility.

The parameters presented in Table 3 are the starting points of the GA. These parameters were obtained as a result of experimental studies. Establishing the model's starting population is the next phase in the hybrid model that has been presented.

Each gene in the initial population is created together with an input selection for that gene. Gene sequencing in the hybrid model is shown in Fig. 5; NI stands for input values, NHL for hidden layer count, NN for neurons per layer count, activation functions for AV hidden layers, and TV for learning algorithm. Each gene's (G) NI value is randomly chosen from the values listed in Table 2 during the population generation step. The function shown in Eq. 5 was used to confine the NHL value for each gene, which was determined randomly.

$$G_a\mathrm{NHL}_b(z) = \begin{cases} 1 & z < 1 \\ z & 1 \le z \le 12 \\ 12 & z > 12 \end{cases} \tag{5}$$

The $G_a\mathrm{NHL}_b(z)$ given in Eq. 5 represents the NHL value of gene a in b iteration.

The NN value for each gene was determined randomly and constrained by the function presented in Eq. 6.

$$G_a\mathrm{NN}_b(z) = \begin{cases} 1 & z < 1 \\ z & 1 \le z \le 12 \\ 12 & z > 12 \end{cases} \tag{6}$$

The $G_a\mathrm{NN}_b(z)$ given in Eq. 6 represents the NN value of gene a in the b iteration.

The AV value for each gene was determined randomly and constrained by the function presented in Eq. 7.

$$G_a\mathrm{AV}_b(z) = \begin{cases} 1 & z < 1 \\ z & 1 \le z \le 6 \\ 6 & z > 6 \end{cases} \tag{7}$$

The $G_a\mathrm{AV}_b(z)$ given in Eq. 7 represents the AV value of gene a in the b iteration. Activation functions corresponding to the determined $G_aAV_b(z)$ value are shown in Table 4.

The TV value for each gene was determined randomly and constrained by the function presented in Eq. 8.

Table 4 AV parameters and values

AV values	Corresponding activation functions
1	Hardlim
2	Tansig
3	Hardlims
4	Radbas
5	Purelin
6	Logsig

$$G_a\mathrm{TV}_b(z) = \begin{cases} 1 & z < 1 \\ z & 1 \le z \le 15 \\ 15 & z > 15 \end{cases} \tag{8}$$

The $G_a\mathrm{TV}_b(z)$ given in Eq. 8 represents the TV value of gene a in the b iteration. The learning functions corresponding to the determined $G_a\mathrm{TV}_b(z)$ value are shown in Table 5.

The network structure for each gene is generated for the hybrid model's subsequent phase based on the values found in its chromosomes. Each G is created in accordance with the input selection chosen for it, and the training-derived verification data are used to assess each G's performance. The fitness functions provided in Eqs. 9–11 are used to assess the effectiveness of the ANN structure produced with each gene.

$$\mathrm{Tf}(G_k) = \text{Learning-Accuracy}\,(\mathrm{ANN}_k) \tag{9}$$

$$\mathrm{Vf}(G_k) = \text{Test-Accuracy}\,(\mathrm{ANN}_k) \tag{10}$$

$$\mathrm{Ff}(G_k) = \mathrm{Max}(\mathrm{Tf}(G_k) + \mathrm{Vf}(G_k)) \tag{11}$$

In Eq. 9 the $\mathrm{Tf}(G_k)$ value represents the Learning-Accuracy fitness value of the k gene and the ANN_k value refers the ANN generated for the k gene. In Eq. 10, the $\mathrm{Vf}(G_k)$ value shows the Test-Accuracy fitness value of the k gene. In Eq. 11, $\mathrm{Ff}(G_k)$ value represents the fitness value of gene k.

Table 5 TV parameters and values

TV values	Corresponding learning functions
1	Trainbfg
2	Traincgb
3	Trainbr
4	Traincgf
5	Traincgp
6	Traingd
7	Traingda
8	Traingdm
9	Traingdx
10	Trainoss
11	Trainrp
12	Trainscg
13	Trainb
14	Trainc
15	Trainr

Table 6 The most successful gene parameters and values

Network parameters	Values
Input (9)	9–1–4–2–5–8–7–3–6
NHL	1
NN	6
AV	Logsig
NN (output layer)	1
AV (output layer)	Purelin
TV	Trainrp

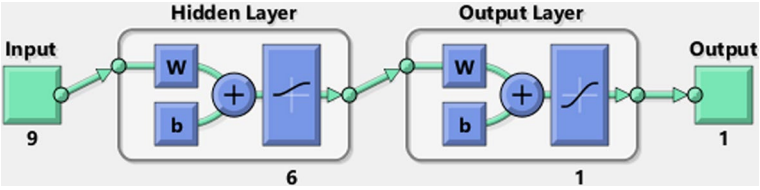


Fig. 6 ANN structure created with the most successful gene

The hybrid model's subsequent stage involved selecting among the population's genes. The roulette wheel approach has been used in the selection process because it generates successful outcomes based on the probabilities of the people who are suited for the solution [36]. Following evaluation of each G's performance in the hybrid model, their cumulative probabilities were computed, and genes were chosen based on the roulette value generated at random.

Crossover between genes in the population comes next after calculating the fitness functions in the hybrid model. The timing of the gene crossover was randomly determined. On the chosen random point, a single-point crossover operation was carried out.

The hybrid model's crossover operation was followed by the application of the mutation process to the population's genes. Each chromosome in each G was assigned a random number between 0 and 1, and the mutation rate shown in Table 3 was used to determine whether the chromosome would change. The chromosomes that were going to be altered received random additions (integer values between −5 and +5). The constraint functions shown in Eqs. 5–8 are applied at this stage.

In order to assist the population move to more promising fields that might contain optimality, elitism was utilized in each iteration [37]. The gene with the highest fitness value was chosen, as can be seen in Table 3, after the fitness values of each gene in the hybrid model that underwent 100 iterations were assessed using the fitness functions provided in Eqs. 9–11. Table 6 shows the ANN network parameters for the gene that produced the highest fitness value.

The network structure established with the network parameters of the most successful gene presented in Table 6 is shown in Fig. 6.

The weights of the data entering the neuron cells and the bias values within the neuron cells are additional factors influencing the network performance in the network topology shown in Fig. 6. Table 7 displays these values that were applied to the network using the hybrid model.

The parameters of the most successful gene derived using the hybrid model and employed in the network are represented by the weights and bias values in Table 7. The learning function and back propagation in the most successful gene were used to determine these values. To repeat the experiment, control the accuracy value achieved, and transfer it to the application, it is crucial to use the network parameters produced with the hybrid model and shown in Tables 6 and 7, which are in the most successful gene.

4 The experimental results and discussıon

In experimental trials for breast cancer detection and prediction, weighted KNN, decision tree, linear SVM, and the presented hybrid model were each used individually. Figure 4 displays the confusion matrix for each model.

According to the confusion matrices presented in Fig. 7, the most successful classifier is the presented hybrid model, which correctly evaluated all test data and correctly predicted all diagnoses.

The ROC curves for each classifier model are shown in Fig. 8. ROC curves are one of the most crucial factors in assessing the effectiveness of models in ML [38]. The FP ratio and TP ratio are represented by the ROC curve's horizontal and vertical axes, respectively. The AUC is the region below this curve [39]. The AUC area of that model is where it performs best, and the optimal value for AUC is 1. As compared to other ML techniques, the hybrid model given in Fig. 8 performed the best.

Precision–recall–F-measure: According to Eqs. 1–4, the accuracy values of the ML approaches employed to classify breast cancer are shown in Table 8.

The values shown in Table 8 show that the suggested hybrid model outperformed other classifiers in terms of success rates. Table 9 displays the results of comparing the suggested model with research done using the same dataset in the literature.

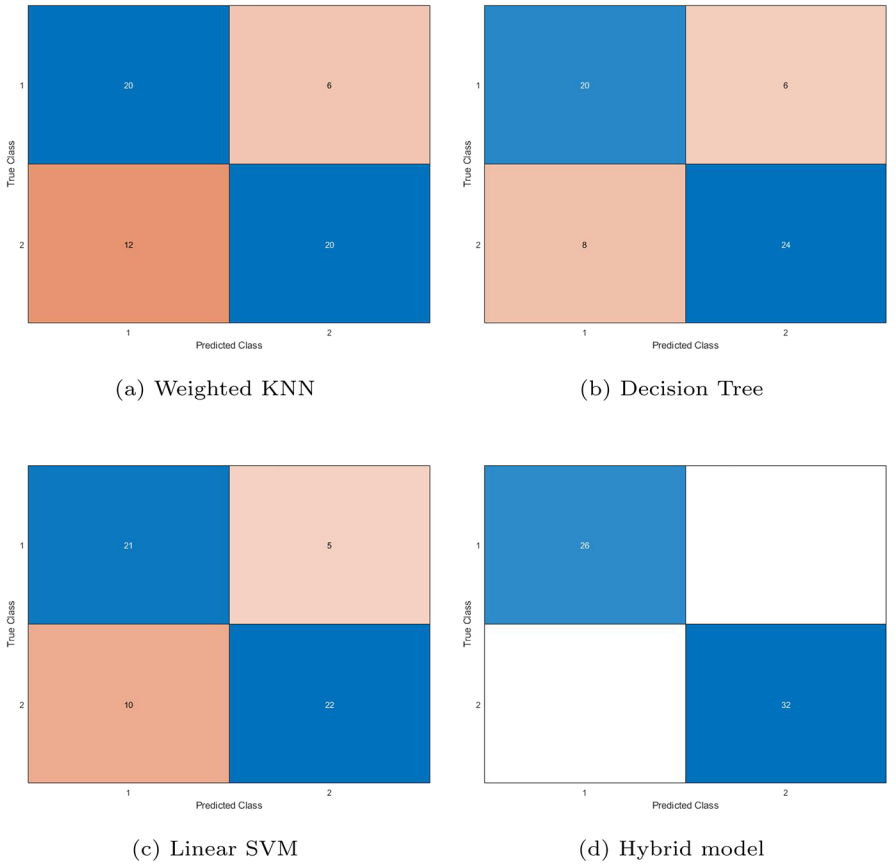
As can be seen in Table 9, the research using the proposed model yielded better outcomes than other studies using the same dataset in the literature.

5 Application

In this part of the research, an iOS–Android-based application was created for the detection and prediction of breast cancer utilizing the experimental data. The values of the parameters listed in Table 6 were used to generate the ANN structure in the built application. The network topology has a single hidden layer, and the hidden layer uses the logsig activation function. Equation 12 illustrates the mathematical expression of the Logsig activation function.

Table 7 Weights and bias values found in the network for the most successful gene

Layers	Weights	Neurons					
		N1	N2	N3	N4	N5	N6
Layer 1	w1	0.00852455894846078	−0.0242388351329295	0.00140207181481561	−0.000672517529240154	0.00463057132963516	0.0144631989084254
	w2	0.0770604989862658	−0.0826962628909338	−0.0121144744316970	0.0284383969913968	−0.0763734204892651	−0.321977173134939
	w3	0.420593772386965	−0.231441017916623	−0.0653822335037580	−0.0428287380249977	−0.178919221189972	0.224479051905515
	w4	−0.155305719144869	−0.100797630147008	−0.239282006636218	−0.250290012154336	0.00855471581413810	0.176734798936250
	w5	0.873551617781607	−0.238211361349990	0.262331886419314	0.409084169342414	−3.374520618571423	0.0979888316870841
	w6	0.0576083190501163	0.152587700375004	0.147638380440604	0.0396465945587499	−0.0491846830986832	0.739942617852241
	w7	−0.276572539383677	0.180097451604797	−0.0121084827885899	−0.0388745115728514	0.0547492042784045	−0.107075465381101
	w8	−0.0551037829888500	0.0303310799922152	0.0591796228381607	0.0258164806763692	−0.0123459962722722	0.0253996773321372
	w9	−0.253942072684184	0.0869907826471019	0.0383218190969294	−0.0716459167103211	−0.00950564227726438	0.0654579914977785
	Bias	8.089909178870473	1.068310074322605	−0.757504432132346	6.638407660986081	2.869955175635420	−0.0598532790782398
Layer 2	w1	1.313641533212121					
	w2	−1.597705372053851					
	w3	0.516286743454256					
	w4	1.497895649512600					
	w5	−6.719312469957271					
	w6	0.849638762870286					
	Bias	0.586480241120817					



(a) Weighted KNN (b) Decision Tree

(c) Linear SVM (d) Hybrid model

Fig. 7 Confusion matrices of the models used

$$\text{logsig}\,(a) = \frac{1}{1+e^{a}} \tag{12}$$

In the developed application, purelin activation function is used in the output layer. The mathematical expression of the Purelin activation function is shown in Eq. 13.

$$\text{purelin}\,(a) = a \tag{13}$$

With the network parameters shown in Tables 6 and 7, the ANN structure was programmed in Android Studio. The created application first arranges the input data as shown in Table 6. The neurons in the hidden layer receive the sorted input data. Considering the weight and bias values in Table 7, the sending operation is conducted. Each neuron's activation function, shown in Eq. 12, is used for filtering the results that result from this procedure. According to the weight and bias values shown in Table 7, the values from the activation functions are re-added to the neuron cell in the output layer. This gathered value is sent through the purelin function

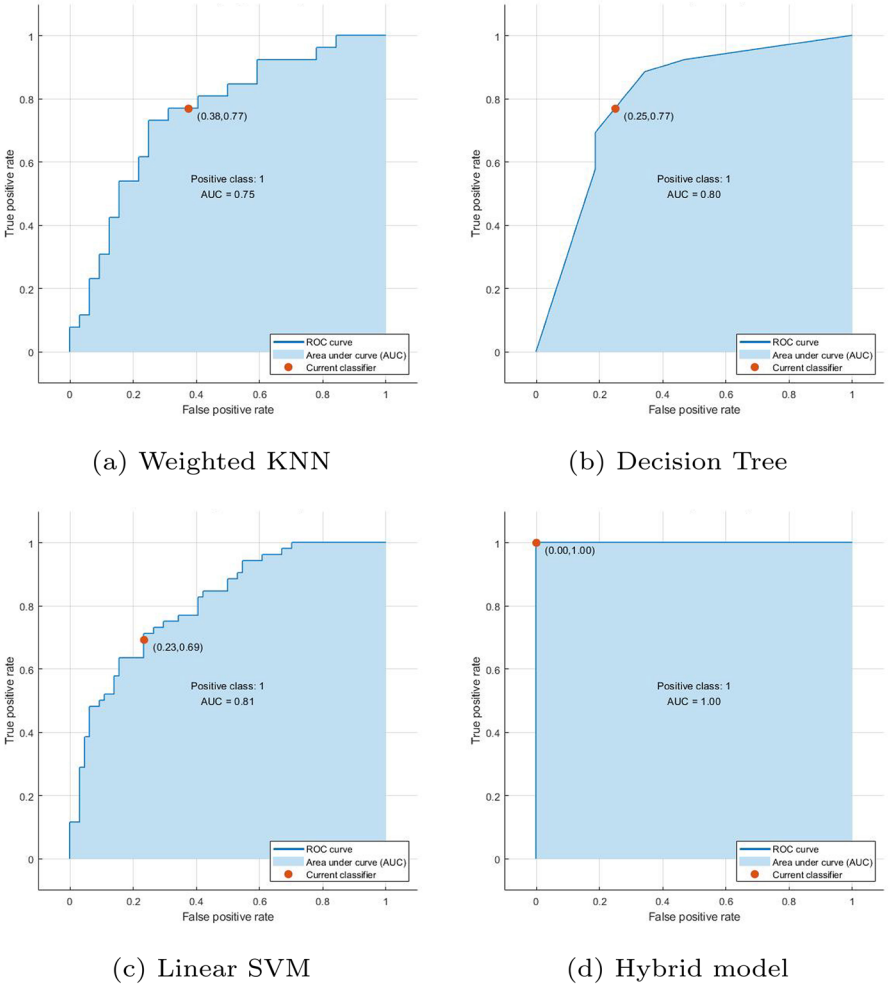


Fig. 8 ROC curves of machine learning methods

Table 8 Experimental results for breast cancer classification

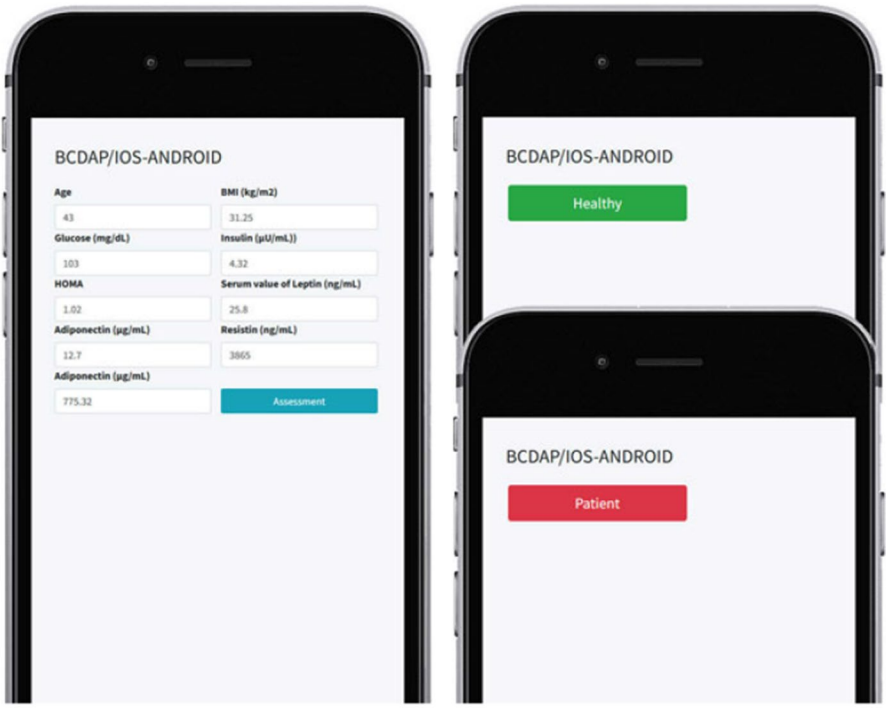
Model	Training	Test	Precision	Recall	F1 score	Accuracy(%)
Weighted KNN	%50	%50	0.63	0.77	0.69	0.69
Decision Tree	%50	%50	0.67	0.77	0.72	0.76
Linear SVM	%50	%50	0.68	0.81	0.74	0.74
Proposed model	%50	%50	1	1	1	100

Table 9 Literature comparison with the proposed model

Authors	Year	Data number	Method	Accuracy (%)
Aslan at al. [11]	2018	116	ELM approach	83.80
Fijri and Rustam [12]	2018	116	SLFCM	89.20
Polat and Senturk [32]	2018	116	Hybrid ML	91.37
Araújo at al. [13]	2019	116	Ornet	81.04
Shuran and Yian [33]	2020	116	PSO-SVM (GP-SVM)	95.65
Yavuz and Eyupoglu [14]	2020	116	PCA-GRNN	97.73
Alshutbi at al. [15]	2022	116	Jaya-SVM	98.21
Proposed method	2023	116	GA-ANN	100

shown in Eq. 13 to produce output information. In terms of breast cancer, a person is healthy if the output value is 0; otherwise, a person is sick if it is 1. Figure 9 illustrates screenshots of the application that was created.

In Fig. 9a, the user provides the input parameters that are employed in the ANN structure. The network structure built by the software designed with modular coding is traversed by the input data that has been received. The user is notified as shown in



(a) Information entry (b) Information interface

Fig. 9 Application screenshots

Fig. 9b with the output. On all smart devices, this designed application is simple to install and use.

6 Conclusion and discuss

Breast cancer is one of the most significant illnesses affecting women worldwide. Early detection of breast cancer and the beginning of therapy have a significant impact on the eventual eradication of the condition.

In this research, a hybrid model for the diagnosis of breast cancer based on GA-ANN is proposed. The Breast Cancer Coimbra Dataset, which is frequently used in the literature, was used as the dataset in the suggested model. The number and order of the input data, the number of hidden layers between the input and output layers, the number of neurons in these layers, the activation functions applied to the neuron cells in the layers, and the learning algorithm of the network are all optimized as a result of the hybrid model that was developed. The investigations conducted thus far on this dataset have produced better results, according to experimental findings.

Along with the hybrid model offered:

- A model with greater accuracy than the research in the literature has been provided using the characteristics in the BCBD data set, which is frequently employed in the detection and diagnosis of breast cancer.
- The number of hidden layers, the number of neuron cells to use in each layer, the activation functions to use in each layer, and the propagation algorithm to use in the network structure that should be established with ANN for the diagnosis and diagnosis of breast cancer were all optimized.
- The network's optimal input data transmission sequence has been determined.
- The model generated the best network models for these specified characteristics, increasing the success rate while also having selected features.
- The model that is being given makes it possible to access these parameters by optimizing every application-specific parameter.

In accordance with these findings, an iOS–Android application was created using the parameters selected by the hybrid model described. To effectively manage this condition, early detection and planning are essential.

Additionally, with regard to the iOS–Android app developed utilizing the data from the hybrid model:

- Doctors would be able to undertake early diagnosis and diagnosis even in rural regions where routine blood analysis may be done.
- Early diagnosis will be made and the disease's death rate will go down as a result of the application that has been developed.
- With specific values, even medical professionals without specialized training can quickly and accurately identify this serious disease in its early stages.
- The use of powerful medications for the patient, which are utilized in the latter stages of the illness, as well as considerable time and expense savings from the

need for these medications and surveillance, will be provided together with early detection and diagnosis.
- Prescanning can be performed quickly with the application.

This application can simply be utilized as a substitute way in health decision support systems.

In light of the findings obtained in future studies, a web-based diagnosis and prediction system can also be developed.

Author Contributions M.A.B. contributed to conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, and supervision.

The authors contributed equally to the work.

Funding There is no funding support for the study.

Declarations

Conflict of interest The authors declare that there is no conflict of interest between them.

Ethical approval The authors declare that no ethical approval is required for this study.

References

1. Santos TBd, Borges AKdM, Ferreira JD, Meira KC, Souza MCd, Guimarães RM, Jomar RT (2022) Prevalence and factors associated to advanced stage breast cancer diagnosis. Ciência & Saúde Coletiva 27:471–482
2. Kayikci S, Khoshgoftaar TM (2023) Breast cancer prediction using gated attentive multimodal deep learning. J Big Data 10(1):1–11
3. Łuczyńska E, Pawlak M, Popiela T, Rudnicki W (2022) The role of ABUS in the diagnosis of breast cancer. J Ultrasonogr 22(89):76–85
4. Elbaiomy M, Akl T, Atwan N, Elsayed AA, Elzaafarany M, Shamaa S (2020) Clinical impact of breast cancer stem cells in metastatic breast cancer patients. J Oncol 2020
5. Lin R-H, Kujabi BK, Chuang C-L, Lin C-S, Chiu C-J (2022) Application of deep learning to construct breast cancer diagnosis model. Appl Sci 12(4):1957
6. Rasool A, Bunterngchit C, Tiejian L, Islam MR, Qu Q, Jiang Q (2022) Improved machine learning-based predictive models for breast cancer diagnosis. Int J Environ Res Public Health 19(6):3211
7. Jayandhi G, Jasmine J, Joans SM (2022) Mammogram learning system for breast cancer diagnosis using deep learning SVM. Comput Syst Sci Eng 40(2):491–503
8. Escorcia-Gutierrez J, Mansour RF, Beleño K, Jiménez-Cabas J, Pérez M, Madera N, Velasquez K (2022) Automated deep learning empowered breast cancer diagnosis using biomedical mammogram images. Comput Mater Continua 71:3–4221
9. Rani S, Kaur M, Kumar M (2022) Recommender system: prediction/diagnosis of breast cancer using hybrid machine learning algorithm. Multimedia Tools Appl 81(7):9939–9948
10. Tarawneh O, Otair M, Husni M, Abuaddous HY, Tarawneh M, Almomani MA (2022) Breast cancer classification using decision tree algorithms. Int J Adv Comput Sci Appl 13(4)
11. Aslan MF, Celik Y, Sabancı K, Durdu A (2018) Breast cancer diagnosis by different machine learning methods using blood analysis data. Int J Intell Syst Appl Eng
12. Fijri AL, Rustam Z (2018) Comparison between fuzzy kernel c-means and sparse learning fuzzy c-means for breast cancer clustering. In: 2018 International Conference on Applied Information Technology and Innovation (ICAITI). IEEE, pp 158–161

13. Silva Araújo VJ, Guimarães AJ, de Campos Souza PV, Rezende TS, Araújo VS (2019) Using resistin, glucose, age and BMI and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer. Mach Learn Knowl Extr 1(1):466–482
14. Yavuz E, Eyupoglu C (2020) An effective approach for breast cancer diagnosis based on routine blood analysis features. Med Biol Eng Comput 58(7):1583–1601
15. Alshutbi M, Li Z, Alrifaey M, Ahmadipour M, Othman MM (2022) A hybrid classifier based on support vector machine and Jaya algorithm for breast cancer classification. Neural Comput Appl 1–13
16. Bülbül MA, Öztürk C (2022) Optimization, modeling and implementation of plant water consumption control using genetic algorithm and artificial neural network in a hybrid structure. Arab J Sci Eng 47(2):2329–2343
17. Bülbül MA, Harirchian E, Işık MF, Aghakouchaki Hosseini SE, Işık E (2022) A hybrid ANN-GA model for an automated rapid vulnerability assessment of existing RC buildings. Appl Sci 12(10):5138
18. Bülbül MA, Öztürk C, Işık MF (2022) Optimization of climatic conditions affecting determination of the amount of water needed by plants in relation to their life cycle with particle swarm optimization, and determining the optimum irrigation schedule. Comput J 65(10):2654–2663
19. Jeyaranjani J, Devaraj D (2022) Improved genetic algorithm for optimal demand response in smart grid. Sustain Comput Inform Syst 35:100710
20. Nahavandi D, Alizadehsani R, Khosravi A, Acharya UR (2022) Application of artificial intelligence in wearable devices: opportunities and challenges. Comput Methods Programs Biomed 213:106541
21. Hasan SSU, Ghani A, Din IU, Almogren A, Altameem A (2022) Iot devices authentication using artificial neural network. Comput Mater Contin 70:3701–3716
22. Dudani SA (1976) The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst Man Cybern 4:325–327
23. Ma Y, Zhao X (2021) Pod: a parallel outlier detection algorithm using weighted KNN. IEEE Access 9:81765–81777
24. Geler Z, Kurbalija V, Ivanović M, Radovanović M (2020) Weighted KNN and constrained elastic distances for time-series classification. Expert Syst Appl 162:113829
25. Dinesh T, Rajendran T (2021) Higher classification of fake political news using decision tree algorithm over Naive Bayes algorithm. Revista Geintec-gestao Inovacao E Tecnologias 11(2):1084–1096
26. Quang-Huy T, Doan PT, Yen NTH, Tran D-T (2021) Shear wave imaging and classification using extended Kalman filter and decision tree algorithm. Math Biosci Eng 18:7631–7647
27. Khan MS, Khan L, Gul N, Amir M, Kim J, Kim SM (2020) Support vector machine-based classification of malicious users in cognitive radio networks. Wirel Commun Mobile Comput 2020
28. Fix E, Hodges J (1951) Discriminatory analysis, nonparametric discrimination: Consistency properties USAF school of aviation medicine, Randolph field. Technical report, Texas, Tech. Report 4
29. Reddy OY, Chatterjee S, Chakraborty AK (2021) Bilayered fault detection and classification scheme for low-voltage dc microgrid with weighted KNN and decision tree. Int J Green Energy 1–11
30. AYIK YZ, Özdemir A, Yavuz U (2007) Lise türü ve lise mezuniyet başarisinin, kazanilan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 10(2):441–454
31. Zhou S, Zhou W (2021) Unified SVM algorithm based on LS-DC loss. Mach Learn 1–28
32. Polat K, Sentürk U (2018) A novel ml approach to prediction of breast cancer: combining of mad normalization, KMC based feature weighting and adaboostm1 classifier. In: 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, pp 1–4
33. Shuran C, Yian L (2020) Breast cancer diagnosis and prediction model based on improved PSO-SVM based on gray relational analysis. In: 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). IEEE, pp 231–234
34. Patrício M, Pereira J, Crisóstomo J, Matafome P, Gomes M, Seiça R, Caramelo F (2018) Using resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer 18(1):1–8
35. Archive.ics.uci.edu: UCI Machine Learning Repository. Last accessed 09 November 2022 (2022). https://archive.ics.uci.edu/ml/datasets.php
36. Wahyuni I, Mahmudy WF (2017) Rainfall prediction in Tengger, Indonesia using hybrid Tsukamoto Fis and genetic algorithm method. J ICT Res Appl 11(1):38–55
37. Li Q, Liu S, Bai Y, He X, Yang X-S (2022) An elitism-based multi-objective evolutionary algorithm for min-cost network disintegration. Knowl-Based Syst 239:107944

38. Tiwari A, Dadhania AV, Ragunathrao VAB, Oliveira ER (2021) Using machine learning to develop a novel Covid-19 vulnerability index (C19VI). Sci Total Environ 773:145650
39. Layer YC, Menzenbach J, Layer YL, Mayr A, Hilbert T, Velten M, Hoeft A, Wittmann M (2021) Validation of the preoperative score to predict postoperative mortality (POSPOM) in Germany. PLoS ONE 16(1):0245841

Authors and Affiliations

Mehmet Akif Bülbül[1]

✉ Mehmet Akif Bülbül
makifbulbul@nevsehir.edu.tr

[1] Faculty of Engineering-Architecture, Computer Engineering, Nevşehir Hacı Bektaş Veli University, Nevşehir 50300, Turkey