Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:39
https://doi.org/10.28919/cmbn/8397
ISSN: 2052-2541

HEALTHCARE PROGNOSIS: GRADILEARN-DRIVEN ELITIST GENETIC ALGORITHM FOR DISEASE PREDICTIONS

B. SAI LAKSHMI, G. GAJENDRAN*

Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, 603203, Tamil nadu, India

Abstract. Hybridization in evolutionary algorithms is gaining traction, boosting convergence speed and solution accuracy-a pivotal research focus. This paper introduces ensemble of GradiLearn with Non-Dominated Sorting Genetic Algorithm-II designed to optimize the neural networks with a focus on dual objectives: enhancing accuracy and minimizing Mean squared error. This paper implements GradiLearn, a robust approach of back propagation with self adaptive learning learning rate. The central concept of the proposed framework involves initiating the population through GradiLearn, rather than relying on random selection. The GradiLearn serves as a tool to optimize weights in cases where weight represents an volatile population parameter. Subsequent to population creation, it evolves with NSGA-II method, namely GRL-NSGA II to produce better generation. The efficacy of GRL-NSGA II is elevated through the enhancement of individuals within the population. This article also implements a non cooperative fitness function for the finest measure called as Accurate Classification Rate (ACR)and Canberra distance-based crowding distance, providing an absolute measure of distance. Experimental results highlight the proposed method's effectiveness in addressing binary and multi-class classification challenges, particularly with imbalanced medical datasets. Through empirical demonstration, the article establishes the models competence in reducing neural network topology while enhancing generalization performance. Comparative analysis with various machine learning models and ensemble methods reinforces the proposed method as a robust classifier, enhancing classification process ability.

*Corresponding author
E-mail address: gajendrg@gmail.com
Received December 14, 2023

Keywords: MLFFN; backpropagation; gradilearn; NSGA II; GRL-NSGA II.

2020 AMS Subject Classification: 68T07, 92B20.

1. PRELIMINARIES

Artificial Intelligence (AI) involves developing computer systems that can perform tasks requiring human intelligence. The goal is to construct machines that can replicate cognitive capabilities and adapt to new knowledge when doing complex tasks. Machine learning (ML), a subset of AI, allows systems to recognize patterns and predict outcomes without the need for explicit programming. The three main types are supervised, unsupervised, and semi-supervised learning, all of which try to improve computer performance through experience. Algorithms such as Decision trees, Random forests, Logistic regression, Support vector machines, and Neural networks are used to handle tasks such as classification, regression, and clustering. Neural networks (NN), which process information through interconnected nodes, play an important part in machine learning tasks like as pattern recognition and decision-making [8].

1.1. Multi-layer Feed-Forward Neural Network. An MLFFNN, or machine learning-optimized feed forward neural network, is designed to process information in a unidirectional flow, moving from input through hidden layers to output. This architecture is particularly adept at mastering complex patterns, making it well-suited for tasks like pattern recognition and classification. Techniques implemented in training include Gradient Descent, Back propagation (BP), and Stochastic Gradient Descent. Back propagation (BPN) is a crucial approach for reducing differences between actual and desired output signals by altering synaptic weights. The addition of a bias weight adds another dimension, improving weight adjustment convergence [29]. The BPN, which is an utilized neural network structure has its set of advantages and disadvantages. One of its weaknesses is that it encounters difficulties when dealing with the ”scaling problem ” resulting in decreased performance when faced with high dimensional data. Another challenge is navigating through sparse global minima within spaces. Additionally BP has limitations when it comes to differentiable criteria or node functions which restricts its usage on certain nodes [23]. A layered feed forward network assures that every path from an input node to an output node has the same number of arcs. The nth layer in this network is made up

of nodes that are n arc traversals from an input node. Each node in layer I is connected to all nodes in layer $i+1$ for all layers, and the network is fully connected if each node in layer I is connected to all nodes in layer $i+1$ for all levels. Because of their effective generalization, layered feed forward networks are popular. They are frequently used in conjunction with the BP training algorithm, which determines optimal weights and biases efficiently. A least-squares optimality criterion is commonly used in BP, a form of gradient search. The idea is to compute the gradient of the error with respect to weights by propagating the error backward through the network [21].

1.2. Emergence of Evolutionary algorithms. To tackle BPN's limitations, there has been evolution in optimization algorithms coupled with machine learning. MLFFNN training can be formulated as an optimization problem. Optimization, a fundamental aspect of machine learning, involves finding the best solution or optimal value for a given problem. Most machine learning algorithms focus on building an optimization model and learning parameters from provided data. To alleviate the mentioned drawbacks of BPN, multi-objective optimization is employed to enhance performance. Real-world problems often involve non-linear and multi-objective nature, making this approach particularly relevant.

Multi-objective optimization (MOO) deals with problems having multiple objectives to optimize. A MOO problem consists of q decision variables, p objective functions (to be minimized or maximized), and constraints (m inequalities and n equalities). The goal is to find optimal solutions considering conflicting objectives.

The goal of optimization is

$$\begin{aligned} Min\ f(x) &= (f_1(x), f_2(x), f_3(x)...f_n(x))^T,\ p \geq 2 \\ subject\ to\ g_i(x) &\leq 0,\ i = 1,2,...,m \\ h_j(x) &= 0, j = 1,2,...,n \end{aligned} \tag{1}$$

Such as m, n is the inequality and equality constraints respectively where $x = (x_1, x_2, x_3, ..., x_q)$ is a q-dimensional decision vectors $x \subseteq \mathscr{R}^q$, f is the n-dimensional objective vector in $\mathscr{R}$. In the era of vast data, evolutionary optimization algorithms significantly impact the adoption of machine learning models. Various efficient optimization techniques have been proposed to

advance machine learning, notably improving effectiveness and efficiency. The Evolutionary Algorithm (EA) mimics nature's evolutionary process, working with a population of solutions in each iteration. This is particularly beneficial for Multi-Objective Optimization Problems (MOOP), where EAs can yield multiple optimal solutions in one simulation run. In contrast to back propagation, genetic algorithms typically do not face the same scaling issues. One key factor is their tendency to consistently enhance the current best candidate. This is achieved by retaining the top-performing individual within the population while exploring for superior candidates. Additionally, genetic algorithms are generally immune to the challenges posed by local minima. The mutation and crossover operators enable them to navigate from one valley to another, potentially reaching a lower point without the hindrance of descending directly into a valley. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is a popular example, emphasizing the need of discerning diverse solutions with varying objective values in Multi-Objective Optimization Problems (MOOP) [5].

1.2.1. *Basic Structure of NSGA II.* NSGA-II enhances the NSGA algorithm by incorporating an improved sorting algorithm, elitism, and eliminating niche requirements. The process begins with a randomly initialized population, which is sorted using non-domination to create hierarchical fronts. Crowding distance values, determined by proximity in the objective space, guide the selection process. Recombination and mutation operators are then applied to generate offspring from parent solutions. This iterative cycle persists until a termination criterion is satisfied. In each generation, the offspring and current population are merged. Elitism ensures the preservation of the best solutions, and non-dominated sorting shapes the new generation. If the population surpasses the current size, solutions are chosen based on crowding distance. This entire process repeats for subsequent generations [22]. The paper is organized into 6 sections. Section 2 highlights relevant papers on multi-objective neural network training, offering background information on selected MLFFNN training methods. In Section 3, the proposed GRL-NSGA II for MLFFNN generalization is detailed, including objectives and constraints. Sections 4 focus on the dedicated discussion of the multi-objective optimization problem, an ensemble approach of BPN and NSGA-II with some modifications. Section 5 outlines experimental results before the conclusion.

2. Related Works

Evolutionary computing is expansively applied to neural networks, facilitating the optimization of crucial elements like activation functions, hyper-parameters (including learning rates), architectural aspects (comprising layer count and neuron distribution), and even the learning rules. In supervised learning, machine learning often involves multiple objectives. MLFNN training is an optimization problem where defining the right objective function is crucial for optimal weight selection. Once the optimization model is formulated, various algorithms can be used, from classical methods like back-propagation (BP) [24] to advanced approaches like genetic algorithms (GA) [7]. In literature, numerous MLFNN training approaches propose using mean square error (MSE) as the optimization model's objective function. However, solely minimizing training error may lead to over fitting. Some works introduce constraints to control network connections and neurons [3]. The process yields an optimized neural network architecture. The learning model must not only achieve favourable approximation performance on the training data but also on unseen data from the same problem. Addressing the limitations of classical methods in handling multi-objectivity, a new approach was explored: the Pareto-based approach [10]. This approach views the objective function as a vector [6]. More papers are adopting a multi-objective approach for MLPNN learning. The initial model in this context, proposed by G. Liu and V. Kadirkamanathan [15], considers three different performance indices as objectives for the multi-objective model. Additionally, Ricardo H. C. Takahashi et al., [1] balanced the training error (MSE) and connection weights during training to enhance MLP generalization. K. Senhaji et al. propose a multi-objective modelling problem with two objectives: accuracy and complexity, given the inherently multi-objective nature of the learning problem. The learning task involves simultaneously minimizing both objectives using NSGAII (Non-dominated Sorting Genetic Algorithm II) as a solver, based on the Pareto domination concept [27].

3. MAIN RESULTS

3.1. Proposed Conglomerate of GradiLearn with NSGA-II. The training of MLFFNN is approached as an optimization problem, with a focus on defining an objective function to minimize MSE and improve accuracy. This article introduces GRL-NSGA-II, a novel approach inspired by existing hybrid models, aiming to overcome their limitations. The proposed ensemble model, combining GradiLearn with NSGA-II (GRL-NSGA II), merges a learning rate-adapted back propagation from GradiLearn with the genetic algorithm NSGA II. GradiLearn incorporates a self-adaptive learning rate in back propagation, and the fitness function is designed to enhance classification accuracy. In contrast to NSGA II's random population initialization, our approach uses GradiLearn to create the initial population. Following population creation, the model progresses through the steps of NSGA II. Building an ensemble model like GRL-NSGA-II presents technical challenges, including optimizing hyper parameters and defining the network architecture. The primary focus lies in defining an appropriate objective function for (1), as it effectively encapsulates the weight selections. Let us elaborate in the section 3.2 as follows:

3.2. Objective Function. When modelling MLFFNN, minimizing Mean Squared Error (MSE) is a primary objective. It serves as a significant function by steering the training process towards attaining greater accuracy and diminishing the discrepancies between predicted and actual values.

In ANN, a given set of inputs $\bar{x} \in \mathscr{R}^n$ has a corresponding label value $y \in \mathscr{R}$, the collection P comprising all possible compact set of patterns of variation $(\bar{x}, y)$.

For an initial value $\bar{v} \in \mathscr{R}^{(n+1)h+h+1}$, an ANN with a hidden layer having $h+1$ neurons on learning set Q_l is defined as $N(\bar{v}, \bar{x})$ and primary objective of modelling is to discover $\bar{v^*} \in \mathscr{R}^{(n+1)h+h+1}$ such that $N(\bar{v^*}, \bar{x}) \cong y$.

In general, the sequence $\{N(\bar{v_i}, \bar{x_p})/p = 1,2,...,P\}_{i=1}^{\infty}$ converges to $\{N(\bar{v^*}, \bar{x_p})/p = 1,2,...,P\}$, where P is the number of elements in Q_l.

$$f_I = \frac{1}{2P} \sum_{p=1}^{P} \{y_p - N(\bar{v_i}, \bar{x_p})\}^2 \tag{2}$$

here $p = 1,2,...,P$ and also equation (2) is the residue between y_p, the actual output and $N(\bar{v_i},\bar{x_p})$, the predicted output at v. In general, the sequence $\{N(\bar{v_i},\bar{x_p})/p = 1,2,...,P\}_{i=1}^{\infty}$ converges to $\{N(\bar{v^*},\bar{x_p})/p = 1,2,...,P\}$ where P is the number of elements in Q_l.

Whenever the network system use the class value $y \in \mathscr{R}$ for each input $\bar{x} \in \mathscr{R}^n$, then the model $\{N(\bar{v_i},\bar{x_p})/p = 1,2,...,P\}$ suffers generalization. Inherently, using MSE (2) alone as an objective function fails to produce the most optimal results $N(\bar{v^*},\bar{x_p})/p = 1,2,...,P$. For any real world problem, minimizing the prediction error is very much important. By incorporating Directional change statistic D_{stat}, the second objective has been proposed aims to minimize MSE and improve accuracy.

$$f_2 = \frac{1}{2P}\sum_{p=1}^{P} a_p * 100 \tag{3}$$

where

$$a_p = \begin{cases} 1 \; if \; \{(y_{p+1} - y_p) * (N(\bar{v_{i+1}},\bar{x_p}) - N(\bar{v_i},\bar{x_p}))\} \geqslant 0 \\ 0 \; otherwise \end{cases}$$

Above equation (3), $min(f_2)$ is used to measure how well a predicted output $N(\bar{v_i},\bar{x_p})$ is closer to actual output y_p. Here P is number of patterns in the neural network. Generally, the value of f_2 lies in between 0 and 1. $f_2 = 0$ means that $y_p = N(\bar{v_i},\bar{x_p})$ for all observations and there is a perfect fit and $f_2 = 1$ means that the performance is under fit.

3.2.1. *Proposed non cooperative objective function.* This proposed work utilizes the Accurate Classification Rate (ACR) as the finest measure for the fitness function in obtaining classification models.

The ACR measure is given in the following expression:

$$ACR = \frac{1}{2P}\sum_{i=1}^{N} P_{ii} \tag{4}$$

where in (4), N is the number of classes, P is the number of patterns in the dataset, and P_{ii} is the number of patterns from the i^{th} class that are correctly classified in that class.

3.2.2. *To define inequality constraints.* In the context of learning in Back propagation Neural Network (BPN), the process can face challenges of getting trapped in local minima, particularly when the determinant value of the Jacobian matrix linked to the outputs of neurons in the hidden or output layers nears zero. To overcome this issue, a suggested optimization approach entails choosing the Jacobian matrix through the imposition of an inequality constraint, as outlined below:

$$\zeta_k(J) = \begin{cases} 1 \; if \; \|J\| \geq d \\ 0 \; if \; \|J\| \leq d \end{cases} ,$$

such that c is the constrained range of interval, $0.1 \leq d \leq 0.9$, k =1,2-layer label and J is a Jacobian matrix of gradients of weight vector, where the criterion function $minf = (f_1, f_2)$ depends on $\bar{v} = (vih, vho)$ or v_{ij}

To avoid the total elimination of updating the weight, at least one neuron to be kept in the k^{th} layer. by using above equation the constraint for equation 1 is defined as

$$\sum \zeta_k(J) \geq 1 \tag{5}$$

So defined objective function (2) and (3) on multi objective optimization problem (1) with the inequality constraint (5).

The equation (1) Multi objective MLFFN training model is defined as follows:

$$\begin{cases} \min f_1(v, \eta, \zeta) \\ \min f_2(v, \eta, \zeta) \\ \text{subject to} \sum \zeta(J) \geq 1; \; k = 1,2 \; layer \; label \\ v = \left(v_{ij}^k\right); \; k = \{0,1\}; \quad 1 \leqslant i \leqslant \{n_0, P\} \; ; \; 1 \leqslant j \leqslant \{P, n\} \\ \eta = (\eta_i); \quad 1 \leqslant i \leqslant P \end{cases} \tag{6}$$

4. A HYBRID OF GRADILEARN WITH NSGA-II

Our proposed model begins with population initialization using proposed GradiLearn approach, tackling the challenge of creating an effective and optimal initial population. The

GradiLearn approach is one which combines a self-adaptive learning rate to enhance the performance of traditional back propagation.

4.1. Adaptive GradiLearn based Initial Population. NSGA-II excels in solving complex multi objective optimization problems. In the realm of evolutionary computation, population scaling is crucial, a subset of solutions in the current generation. Random population generation, considering factors like search space, fitness function, diversity, and more, can boost solutions, especially with a larger population. However, initializing based on conventional BPN has drawbacks. Our approach balances population size for better solutions, providing the algorithm with an optimal number of chromosomes, each encoding a potential problem solution. The general form of Population will be

$$P_i = \{v_i / v_i \in R\} \tag{7}$$

In (7) weight vector v_i serves as a pivotal element in the optimization process through GRL-NSGA II. Initially, this vector is conceived as an ensemble of components drawn randomly from the population associated with the target population elements.

To initiate training of a model, a random weight vector v_1 is selected from the neural network. Subsequently, the initial weights undergo modification under proposed self adaptive learning rate. In this process, v_i is regarded as a fundamental hyper parameter, and its values are refined through successive iterations of GradiLearn approach.

The initial population parameters are now explicitly defined, selected from the specified range of intervals below.

$$P_1 = \{v_1 / v_i \in R^n, i = 1, 2, ...n\}$$

The slope of y_p layer is almost parallel to x_p and least at every point of the interval ranges from $[-\infty, \infty]$, where initial learning of MLFFNN gets slow. It is important to initialize some neurons in the proposed interval x_p ranges from $-3.4 \leq x_p \leq 3.4$.

For any real number $r > 1$, choice of weight vector (v_1) lies between $\left[-q\sqrt{\frac{l}{(m+1)max\|x^u - x^v\|_2}}, q\sqrt{\frac{l}{(m+1)max\|x^u - x^v\|_2}}\right]$ where $(u, v) \in P$ and $x^u, x^v \in x_p$, where l is user adaptive fixed constant and fixed value for $q = 6.8$.

This standard approach applies to all neurons in the MLFFN. It entails using an analytical presentation to initialize the weight vectors depending on their corresponding input vectors, transforming the MLFFN architecture into a mapping or function, also known as mapping representation. The major goal of this mapping representation is to ensure that the MLFFN can generate output values that fall inside the specified desired output interval.

For an initial weight $\bar{v_1} \in \mathscr{R}^{(n+1)h+h+1}$ and the input vector $\bar{x} \in Q_l$, and $\left\{N(\bar{v_1}, \bar{x_p})/p = 1,2,\ldots,P\right\}$ a NN with a hidden layer having $h+1$ hidden neurons on training set from input vector Q_l and it is a first population. The updation of weight vector is dependent on a learning rate that transports learning information to weight encoded in network and has a positive impact on producing superior generation. For all proposed initialized weight components v_0, the weight updated rule for this model is formally given by

$$v_{i+1} = v_i - \eta_{i-1}\Delta v_i \tag{8}$$

where in (8), Δv_i is the rate of change of f_1 with respect to f_i where $\Delta v = \partial f_1(v)/\partial v$. The updation of weight rule is sensitive to the hyper parameter η, is the learning rate and our proposed self adaptive learning rate is discussed below: To find the next generation, the second population weight vector v_2 is generated using initial learning rate η_0 on conventional BPN.

4.1.1. *Self-Adaptive learning rate.* In the process of generating the subsequent population weight vector, this article introduces an enhanced variant known as GradiLearn. This modification involves the incorporation of an Self-Adaptive learning rate , denoted as η (the learning parameter that imparts information to the next generation). The proposed adaptation of the learning parameter is founded on its inherent capacity to acquire knowledge.

In gradient descent learning algorithms like error back propagation, the learning rate parameter plays a crucial role in influencing generalization accuracy. To enhance generalization accuracy, especially in dealing with large and intricate problems, a noteworthy strategy involves diminishing the learning rate beyond the threshold that yields the fastest convergence. Surprisingly, even with a relatively modest learning rate, optimal generalization accuracy can be attained with a minimal number of epochs.

The dynamics of the population structure can be significantly updated through the process of learning. Therefore, we expound on the concept of learning. In this context, each weight vector encapsulates its unique information, and this is denoted as the learning rate. Selecting the optimal learning rate poses a challenge, mainly due to the non-monotonic nature of the process. To ensure that all weight components minimize the error function through the on-line learning method, our proposed model for the learning rate is outlined as follows:

Let $\left\{N(\bar{v_i},\bar{x_p})/p=1,2,\ldots,P\right\}$ be the set of NN corresponding to epochs, where $\{N(\bar{v^*},\bar{x_p})/p=1,2,...,P\}$ is updated network of proposed model.

Consider AEA_t and AEB_t as the absolute errors of all patterns generated by $\left\{N(\bar{v_i},\bar{x_p})/p=1,2,\ldots,P\right\}$ after and before learning, respectively. But due to non-monotonic nature, the NN may diverges. To circumvent this issue, if $AEB_t < AEA_t$, then AEA_t should not surpass $3AEB_t$; otherwise, proposed GradiLearn stops the learning process. In light of this, considering an initial learning rate η_0 close to zero, the proposed learning rate can be expressed as:

$$\eta(t+1)=\begin{cases}0 \quad if\ \left\{N(\bar{v_i},\bar{x_p})\right\}_{i=1}^{\infty}\ converges\ to\ N(\bar{v^*},\bar{x_p}) \\ \eta(t)+b+r\frac{AEA_t}{3*AEB_t}\ \ if\ 0<\frac{AEA_t}{3*AEB_t}<\frac{1}{3} \\ \eta(t)-b-r\frac{AEA_t}{3*AEB_t}\ \ if\ \frac{AEA_t}{3*AEB_t}\geqslant\frac{1}{3}\end{cases} \tag{9}$$

The learning process commences with an initial learning rate, denoted as $\eta_0 = 0.00009$. In the learning phase of real-world problems, our devised model dynamically tunes the learning rate using the equation with parameters $b = 0.001$ and $r = 0.001$. This adaptive approach facilitates the algorithm in efficiently minimizing errors. The suggested learning rate leverages information by assessing the learning capability of GradiLearn. Throughout the learning process, it functions as a self-regulating and adaptive learning rate, aiming to control the non-monotonic behaviour of GradiLearn to the maximum extent[25].

Creating population using equation (8) and (9) tailored to the demands of the real-world problem at hand. The size of this population is contingent upon the elements of the matrix (vectors) associated with the given real-world problem. In our approach, generating population size to accommodate all data samples, capped at 50.

After initializing the weight vector using learning rate, the network initiates GradiLearn, aiming to minimize the mean squared error as its primary objective. Subsequently, it adheres to the conventional NSGA-II procedure to generate improved offspring, with the ultimate goal of converging the mean squared error towards zero and also to generate offspring population solutions. To proceed with NSGA-II algorithm, the initial population is combined with the offspring population.

4.2. Non-dominated sorting. After initial population, utilize non-dominated sorting to select best solutions using concept of Pareto optimality. Pareto optimal is calculated at each epoch using cost function f_1 and f_2.

The solution x_1 dominates the other solution x_2 if it satisfies the condition where, for all objectives, x_1 is at least as good as x_2, and for at least one objective, x_1 is strictly better than x_2:

$$f_i(x_1) \leq f_i(x_2) \; for \; all \; i = 1,2,\ldots,n$$

$$f_i(x_1) \leq f_i(x_2) \; for \; atleast \; one \; i = 1,2,\ldots,n$$

Where n denoting the number of objective functions and $f_i(x)$ representing the i^{th} value of an objective function for decision vector x, the relation is such that x_1 dominates x_2. A solution x from the solution space (i) is not dominated by any other feasible solutions, it is called a Pareto optimal solution. The collection of all Pareto optimal solutions is denoted as a Pareto set, and the corresponding objective vector defines a Pareto front. Then $\left\{N(\bar{v_i},\bar{x_p})/p = 1,2,...,P\right\}$ converges to $\left\{N(\bar{v^*},\bar{x_p})/p = 1,2,...,P\right\}$ using Pareto optimality.

4.3. Canbera distance based Crowding Distance Algorithm (CB-Dist based Crowding Distance Algorithm). Perform non-dominated sorting on the population, assigning a rank value to each network based on its non-dominated level. Furthermore, this paper introduces an innovative method for computing the Crowding Distance measure based on CB-Dist for each individual. The CB-Dist based Crowding Distance of each point in the front F_r is determined using the following algorithm:

Step 1: Assign 0 to the corresponding CB-Dist Crowding Distance, for every solution in the set F_r.

Step 2: Arrange the set in descending order of the worse adjustment of f_m, for each objective function $f_m, m = 1,2,\ldots M$.

Step 3: Assign an infinite distance to the boundary solutions $D^r(1) = D^r(l) = \infty(1)$ is the first solution and l is the last one in the front F_r(after sorting). Then, for all other solutions $i = 2,...,l-1$ Assign:

$$D^r(i) = D^r(i) + \frac{\mid f_m^r(i+1) - f_m^r(i-1) \mid}{\mid f_m^{max} \mid - \mid f_m^{min} \mid} \tag{10}$$

Such as:

F_r The r^{th} Pareto set;

f_m^r (i+1) is the m^{th} objective function value of the (i+1) solution in the set F_r;

f_m^r (i-1) is the m^{th} objective function value of the (i-1) solution in the set F_r

f_m^{max} is the maximum value of the m^{th} objective;

f_m^{min} is the maximum value of the m^{th} objective;

The aggregate CB-Dist based Crowding Distance (10), denoted as $D^r(i)$, is the summation of the crowding distances of solution i concerning each objective.

4.4. Crowded Tournament Selection Operator. To determine candidates for the next generation, the NSGA-II uses tournament selection based on crowding distance criteria. Non-domination rank i_rank and crowding distance i are used to evaluate each member in the population. It can be defined as follows:

$$f(i_{rank} \leq j_{rank}) or (i_{rank} = j_{rank}) and (i_{cr} \leq j_{cr})$$

In the event that two randomly selected solutions, i and j, share the same non-dominated rank $(r_i = r_j)$, the one with the higher crowding distance value is selected. If their crowding distances are identical, the choice is determined by rank, giving preference to the solution with the lower rank. This method safeguards population diversity and mitigates premature convergence.

Following with tournament selection, the optimal candidate solution is chosen, and genetic operators such as crossover and mutation are subsequently applied.

4.5. Genetic operators. The offspring cohort, which results from the parental selection process, lays the framework for the creation of a new group of individuals. Furthermore, the offspring population is divided into two segments, each of which is subjected to either crossover or mutation operations. There is an extensive number of created crossover and mutation procedures in the literature, and this article focuses on the operators that are specifically geared to handle our unique problem.

4.5.1. *Crossover.* The proposed method employs a crossover scheme inspired by the binary crossover namely simulated binary crossover(SBX crossover) observed in nature. This technique is applied individually to each variable.

Simulated Binary Crossover. To calculate the offspring the formulation is as follows:

$$\mathscr{C}_{1,k} = 0.5(1+\alpha_k)f_{1,k} + (1-\alpha_k)f_{2,k}$$

$$\mathscr{C}_{2,k} = (1+\alpha_k)f_{2,k} + 0.5(1-\alpha_k)f_{1,k}$$

Such as $f_{i,k}$ is the selected parent, $\mathscr{C}_{i,k}$ is the i_{th} child with k_{th} component. To calculate α_k using below formulation:

$$\alpha_k = \begin{cases} (2\mu)^{\frac{1}{r+1}}, \ if \ r < 0.5 \\ (\frac{1}{2(1-\mu)})^{\frac{1}{r+1}} \ , \ otherwise \end{cases}$$

where r is the index of user defined distribution(positive) and $0 < \mu < 1$ To compute function of α_k, using probability distribution:

$$f(\alpha_k) = \begin{cases} 0.5(r+1)\alpha^r, \ if \ \alpha \leq 0.5 \\ 0.5(r+1)\frac{1}{\alpha^{r+2}} \ , \ otherwise \end{cases}$$

The utilized crossover rate is 0.7. Crossover does not directly enhance the likelihood of discovering the optimal solution within a given set. However, it plays a crucial role in accelerating the convergence speed toward a potential optimal solution during the search process. Considered a convergence operator, crossover acts as a pivotal step in steering the population towards the optimum value.

4.5.2. *Mutation.* Mutation is not ubiquitous in the evolutionary process. When there is a lack of diversity among individuals, the population is reinitialized to avoid premature convergence or stagnation in the search. Only the best individuals are retained in the new population. The mutation technique employed is described as follows

$$f(i) = \mathscr{C}(i) + (\alpha - 0.5) * 14$$

$f(i)$ is the selected parent, $\mathscr{C}(i)$ is the obtained child and α_k is a chosen by a uniform distribution in $[0,1]$ randomly. Implementing a basic mutation involves exchanging values among randomly selected genes. The genes chosen for permutation are determined based on a specified mutation rate (0.4).

4.6. Replacement. NSGA-II utilizes an elitist algorithm, employing a steady-state population replacement approach. This involves combining the offspring population with the current population to fulfil niching requirements. Moreover, a non-dominated ranking strategy is applied to choose the leading N individuals (with N denoting the population size) from this merged population. These selected individuals subsequently serve as the foundation for generating the next generation of the population.

4.7. Stopping criteria. Our algorithm employs the attainment of an optimal number of iterations as a stopping criterion, ensuring the best convergence rate.

5. Experimental Results and Discussions

In this section, performed range test for our proposed approach using standard benchmark data sets from the University of California Irvine (UCI) machine-learning repository. We chose 4 classification problems with imbalanced data sets: Breast cancer, Cryotherapy, Heart, Immunotherapy. The data set is utilized to test the ability of the proposed model. Additionally, all the benchmark data sets are normalized to same scale [0,1]. The performance evaluation was carried out in MATLAB 2023 software. The discussion about the data set and training method given below in detail for each data set separately. Table 1 represents overview of used data set characteristics: the number of instances, number of components, class and their imbalance ratio

are mentioned. For model construction and validation, this article employed a neighbourhood-based approach to partition the datasets into feasible ratio. This strategic division aimed to enhance the model's overall performance [18].

TABLE 1. Pre-Processing of Dataset

Data sets	Instances	Components	Class	Imbalance ratio	Threshold	Learning set	Testing set
Breast Cancer	569	30	2	1.68396	0.3982	303	266
Cryotherapy	90	7	2	1.14285	0.5439	48	42
Heart	303	13	2	1.19565	0.5386	230	73
Immunotherapy	90	8	2	3.7368	0.6712	63	27

5.1. Breast cancer Data set. Breast cancer is the world's second largest cause of female mortality. Early detection is critical, and advanced machine learning algorithms are being used by researchers to make accurate predictions. The Wisconsin Breast Cancer Dataset (WBCD) from the UCI repository, with 569 instances and 30 features, served as the training set. It's noteworthy for its completeness with no missing values. Out of 569 cases, 357 are benign, and 212 are malignant, providing a robust dataset for assessing machine learning efficacy in breast cancer prediction[28].

5.1.1. *Performance evaluation of GRL-NSGA II.* After pre-processing of data, Performance of proposed method evaluated by various statistical measures to substantiate our model's stability and it is given in Table 2.

TABLE 2. Performance metrics of proposed GRL-NSGA II

Accuracy	Precision	Recall	Specificity	AUC	F-Score	G-Means	MCC	Kappa	No.of Gen
99.649	99.0566	100	99.4429	99.7215	99.5261	99.7211	0.9925	0.9925	2

The performance of each benchmark dataset and experimental results underscores the robustness and benefits of our proposed model. Table 3 presents the performance of our method in comparison to other established machine learning models and hybrid models.

TABLE 3. Performance comparison of various prediction methods in the Breast cancer Dataset

Method	Accuracy	F-Score	Precision
LR [28]	96	-	-
NB [28]	96.61	-	-
SVM [28]	97	-	-
DT [28]	95.81	-	-
CNN [28]	99	-	-
Voting classifier [28]	98.77	-	-
MLP [28]	98	-	-
RF [28]	97.66	-	-
XGBoost [28]	98.24	-	-
KNN [28]	94.35	-	-
Hybrid crowded Pearson [14]	92.31	-	-
Pearson correlation [14]	92.47	-	-
FCA [14]	98.255	-	-
Hybrid SVR using NSGA II [30]	99.07	99.04	0.97
BSO-CV [13]	99.30	-	-
Proposed GRL-NSGA II	99.649	99.5261	99.0566

The above table represents GRL-NSGA II classifier represents an enhanced iteration of the pre-existing model, designed categorizing unfamiliar observations through a completely innovative approach. Consequently, this paper assessed the effectiveness of GRL-NSGA II in identifying breast cancer using best-performing classifiers. (i) K.M.M. Uddin et al. 2023 (various alforithms) (ii) Abdesslem Layeb., 2023 (Hybrid crowded pearson, Pearson correlation, FCA) (iii) Mohammad Hossein) Zangooei et al., 2014 (Hybrid SVR using NSGA II)(iv) R. Abu Khurma et al. 2023 (BSO-CV). In the context of this comparative analysis, distinct software implementations of these classifiers are factored. Remarkably, GRL-NSGA II outperforms the

aforementioned classifiers. As indicated in the provided table, GRL-NSGA II achieves exceptional results with the highest Accuracy, Precision, and F-measure reaching values of 99.649, 99.5261, and 99.05660 respectively in the Breast cancer dataset.

5.2. Cryotherapy. Warts, non-cancerous tumors on the skin, are often treated with Cryotherapy applying extreme cold to eliminate abnormal tissue. Data from a dermatology clinic in Iran (UCI Machine Learning Repository) covers 90 patients' response to Cryotherapy. Sessions, up to ten with one-week intervals, were recorded. If not cured, other methods were explored. The dataset features seven factors, with the outcome variable being Cryotherapy effectiveness. Of 90 patients, it worked for 48, and for 42, it did not. The response variable distribution is balanced (53.3% positive, 46.7% negative). Success rates differ by gender, with Cryotherapy more effective for males (57.4%) than females (48.8%) [17].

5.2.1. *Performance evaluation of GRL-NSGA II.* After pre-processing of data, Performance of proposed method evaluated by various statistical measures to substantiate our models stability and it is given in Table 4.

TABLE 4. Performance metrics of proposed GRL-NSGA II

Accuracy	Precision	Recall	Specificity	AUC	F-Score	G-Means	MCC	Kappa	No.of Gen
97.778	97.9167	97.9167	97.6190	97.7679	97.9167	97.7677	0.9554	0.9554	4

Performance of the individual benchmark data set and experimental results signifies the robustness and advantage of our proposed model. Table 5 shows the performance of proposed method compared with other existing machine learning models and hybrid models.

TABLE 5. Performance comparison of various prediction methods in the Cryotherapy Dataset

Method	Accuracy	F-Score	AUC score	Precision	Recall
MLP [17]	93.3	93.6	-	91.7	95.7
SVM [17]	92.2	92.3	-	87.5	97.7
CVM [17]	97.8	97.9	-	97.9	97.9
k-NN [17]	93.3	93.8	-	93.8	93.8
BLR[17]	91.1	91.3	-	87.5	95.5
CART[17]	89.22	-	-	88.57	89.79
LDA [17]	86.67	-	-	89.05	84.59
GNC [4]	95.03	-	-	-	-
Pruned FNN [11]	88.64	-	0.89	0.86	0.93
FNN [11]	85.75	-	0.85	0.80	0.90
P-dist TWSVM [26]	94.4444	95	-	-	-
Proposed GRL-NSGA II	97.778	97.9167	97.7679	97.9167	97.9167

The GRL-NSGA II classifier is an advanced evolution of the existing model, engineered to classify unfamiliar observations using a completely novel approach. In this context, we evaluated the efficacy of the GRL-NSGA II in precisely differentiating instances within the Cryotherapy dataset, leveraging high-performing classification techniques. (i) Y. Chen, X. Zhang, Y. Zhuang et al. 2023 (GNC) (ii) G Augusto Junio Guimarães et al., 2019 (Pruned FNN, FNN)(iii) Yashik Singh., 2021 (MLP, SVM, CVM, KNN, BLR) (iv) Md Mamunur Rahman et al., 2019 (CART, LDA) (v) Sai Lakshmi B & G.Gajendran (P-dist TWSVM) 2023. In the scope of this comparative examination, diverse software realizations of these classifiers are taken into account. Notably, when juxtaposed with the classifiers mentioned earlier, GRL-NSGA II exhibits superior performance. As emphasized in the provided table, GBPNSGA-II achieves noteworthy results, achieving Accuracy, F-score, AUC score, Precision, and Recall values of 97.778, 97.9167, 97.76785, 97.9167, and 97.9167 respectively on the Cryotherapy dataset.

5.3. Heart-Cleveland. Cardiovascular disease stands as the primary global cause of death. However, enhancing patient survival rates and reducing fatality rates can be achieved through early diagnosis and prognosis. Unfortunately, the shortage of radiologists and doctors in many countries poses a significant obstacle to early detection. The predominant dataset utilized in numerous research papers is sourced from the University of California, Irvine (UCI) Center for Machine Learning and Intelligent Systems. This dataset comprises four databases from distinct hospitals, each sharing a common set of 14 features but varying in the number of records. Specifically, the Cleveland dataset, one of the four, encompasses 303 instances, with 164 corresponding to normal patients and 139 to those with heart conditions [2].

5.3.1. *Performance evaluation of GRL-NSGA II.* After pre-processing of data, Performance of proposed method evaluated by various statistical measures to substantiate our models stability and it is given in Table 6.

TABLE 6. Performance metrics of proposed GRL-NSGA II

Accuracy	Precision	Recall	Specificity	AUC	F-Score	G-Means	MCC	Kappa	No.of Gen
97.6897	97.5757	98.1707	97.1223	97.6465	97.8723	97.6451	0.9534	0.9534	7

Performance of the individual benchmark data set and experimental results signifies the robustness and advantage of our proposed model.

Table 7 shows the performance of proposed method compared with other existing machine learning models and hybrid models.

TABLE 7. Performance comparison of various prediction methods in the Heart Dataset

Method	Accuracy	Precision	Recall
NB [2]	86.4198	-	-
ANN ensemble [2]	89.01	-	-
DT [2]	85	-	-
KNN [2]	97.4	-	-
SVM & RBF [2]	86.42	-	-
BPNN [16]	96	95	95
ELM-NN [16]	87	-	-
GA-SVM [16]	88	-	-
PSO-FFBP [16]	92	92	93
DT-NN [16]	78	-	78
NSGAII [9]	91.6	-	-
Hybrid Ann-NSGA II [19]	93.2	-	-
P-dist TWSVM [26]	84.2975	-	-
Proposed GRL-NSGA II	97.6897	97.5757	98.1707

The GRL-NSGA II classifier signifies a sophisticated progression from the present model, created to categorize the provided instances using a completely novel approach. In this particular context, we evaluated the efficacy of the GRL-NSGA III in precisely discerning cases within the Heart - Cleveland dataset, utilizing leading classification techniques. (i) Maryam I. Al-Janabi 2018 (NB,ANN ensemble, DT, KNN, SVM & RBF) (ii) Samir Malakar et al., 2015 (BPNN, ELM-NN, GA-SVM, PSO-FFBP, DT-NN) (iii) Ashraf Mohamed Hemeida et al., 2018 (NSGA II) (iv) S Mane et al., (hybrid NSGA II and ANN) 2016. Within the context of this comparative analysis, various software implementations of these classifiers are considered. Remarkably, when compared to the previously mentioned classifiers, GRL-NSGA II demonstrates superior performance. Highlighted in the provided table, GRL-NSGA II attains remarkable

outcomes, with Accuracy, Precision, and Recall values of 97.6897, 97.5757, and 98.1707, respectively, when applied to the Heart - Cleveland dataset.

5.4. Immunotherapy. Human Papillomavirus (HPV) infection triggers the formation of warts, often on the hands and feet, posing challenges in later-stage treatment. Literature review reveals diverse proposed treatments, highlighting a key challenge: tailoring treatments for individual patients. This variability makes it hard to identify optimal approaches for faster recovery and personalized treatment. Authors, in their experimentation, utilized Immunotherapy datasets from UCI, featuring detailed information on 8 attributes for a total of 90 patient instances. These 8 features play a crucial role in understanding and analysing immunotherapeutic responses against HPV [11].

5.4.1. *Performance evaluation of GRL-NSGA II.* After pre-processing of data, Performance of proposed method evaluated by various statistical measures to substantiate our models stability and it is given in Table 8.

TABLE 8. Performance metrics of proposed GRL-NSGA II

Accuracy	Precision	Recall	Specificity	AUC	F-Score	G-Means	MCC	Kappa	No.of Gen
93.333	97.1831	94.5205	88.2353	91.3779	95.8333	91.3239	0.7937	0.7918	6

Performance of the individual benchmark data set and experimental results signifies the robustness and advantage of our proposed model. Table 9 shows the performance of proposed method compared with other existing machine learning models and hybrid models

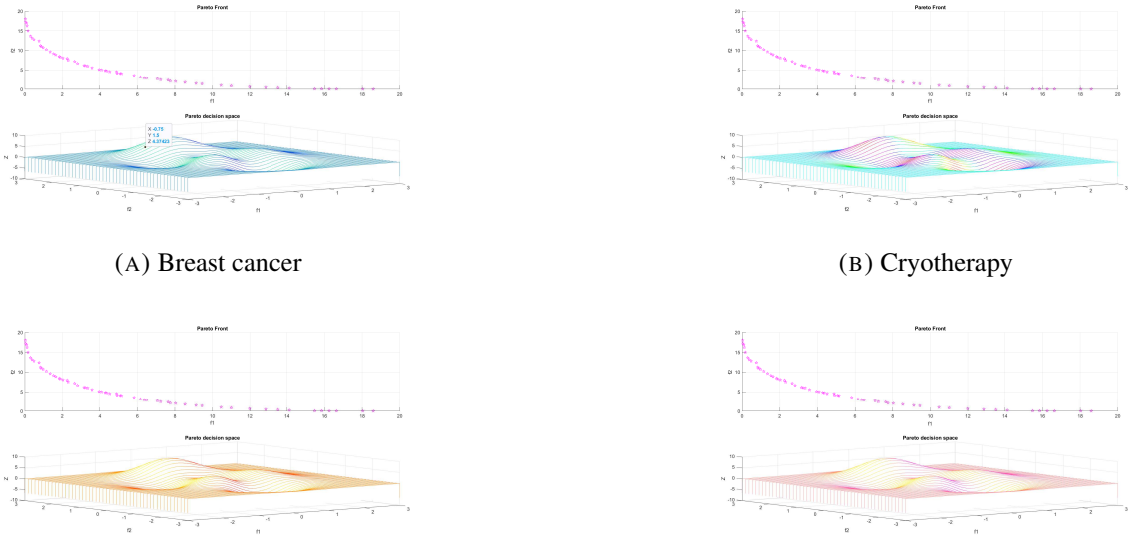
TABLE 9. Performance comparison of various prediction methods in the Immunotherapy Dataset

Method	Accuracy	F-score	AUC score	Precision	Recall
Pruned FNN [11]	84.32	-	0.69	0.97	0.41
FNN [11]	81.91	-	0.65	0.94	0.36
MLP [11]	78.02	-	74	0.88	0.60
J48 [11]	83.92	-	0.71	0.91	0.52
NB [11]	76.67	-	0.69	0.87	0.51
ZR(Zero rule)[11]	79.13	-	0.50	0.50	0.50
RT(Random tree)[11]	81.24	-	0.74	0.94	0.54
SVM [12]	78.9	88.2	-	62.2	78.9
KNN [12]	70	69.6	-	68.2	70
Binary Logistic regression [12]	85.0	91.6	-	100	84.5
Bayes net [12]	85.5556	-	-	-	-
FURIA [12]	82.2	-	-	-	-
RF [12]	86.7	-	-	-	-
C5.0 DT [20]	93.324(Test) & 88.13(Train)	-	-	-	-
Proposed GRL-NSGA II	93.333	95.8333	91.3779	97.1831	94.5205

The GRL-NSGA II classifier signifies a sophisticated progression from the present model, created to categorize the provided instances using a completely novel approach. In this particular context, we evaluated the efficacy of the GRL-NSGA II in precisely discerning cases within the Immunotherapy dataset, utilizing leading classification techniques. (i) G Augusto Junio Guimarães et al., 2019 (Pruned FNN, FNN, MLP, J48, NB, ZR, RT) (ii) Sabita Khatri1 et al., 2018 (SVM, KNN, BLR, BN, FURIA) (iii) Maad M. Mijwil et al., 2021 (C5.0 Decision Tree). Within the context of this comparative analysis, various software implementations of these classifiers are considered. Remarkably, when compared to the previously mentioned classifiers, GRL-NSGA II demonstrates superior performance. Highlighted in the provided table, GRL-NSGA II attains remarkable outcomes, with Accuracy, Precision, and Recall values of 93.333, 95.8333, 91.3779, 97.1831 and 94.5205 respectively when applied to the Immunotherapy dataset.

6. Visual Representation of Proposed GRL-NSGA II

Graphical representations provide a visual overview of complex models, making it easier for individuals to grasp the structure and relationships within the model. This visual clarity is especially important when dealing with intricate and imbalanced data. It is also a fundamental aspect of the modelling process that enhances understanding, communication, and decision-making.



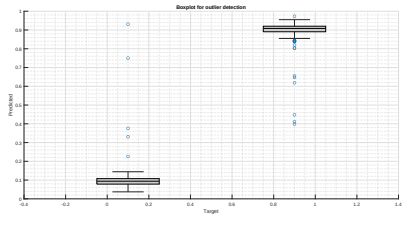
(A) Breast cancer (B) Cryotherapy

(C) Heart-Cleveland (D) Immunotherapy

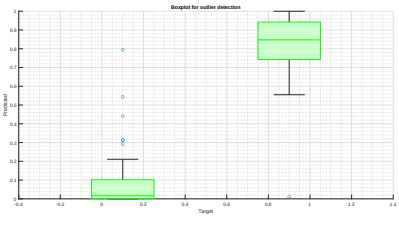
FIGURE 1. Pareto Front defining trade off between conflicting objectives

Figure 1 illustrates the Pareto fronts, delineating the trade-off between conflicting objectives. In the second subplot, the waterfall plot visualizes the region of the decision space where the Pareto optimal solutions are selected. Along the x-axis, it have the first objective, the y-axis represents the second objective, and the z-axis indicates the data number.

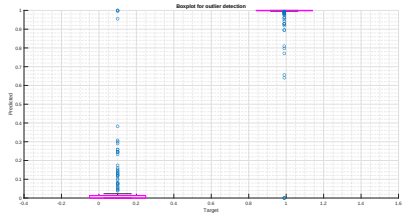
Figure 2 represents box plot, also known as a box-and-whisker plot, is a graphical representation that provides a summary of the distribution of errors between Target (x axis) and the predicted outputs (y-axis). It displays key statistical measures and allows for the identification of potential outliers.



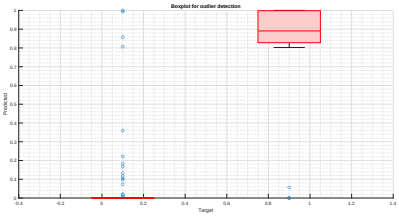
(A) Breast cancer



(B) Cryotherapy



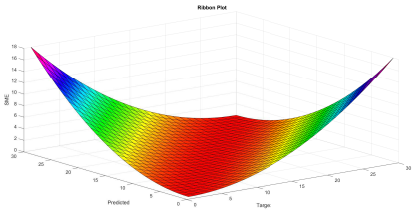
(C) Heart-Cleveland



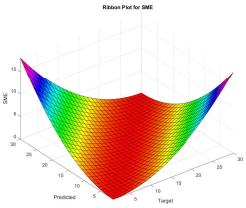
(D) Immunotherapy

FIGURE 2. Misclassification rate: Target vs. Predicted Output

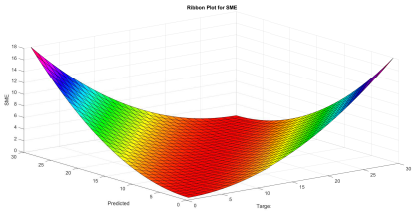
Figure 3 depicts a ribbon plot, serving as an effective representation of uncertainty or variability in the data. In this context, it illustrates the variability of the proposed objective function over time. Additionally, the plot visually conveys the proximity of the objective function to the target output, highlighting efforts to minimize f_1.



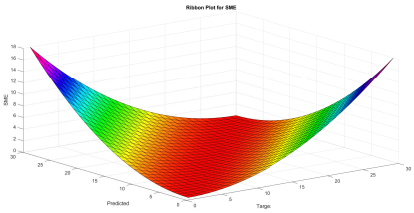
(A) Breast cancer



(B) Cryotherapy

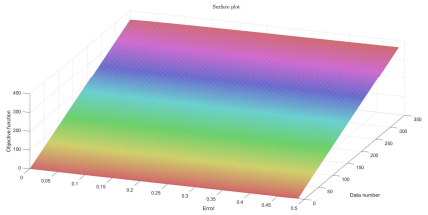


(C) Heart-Cleveland

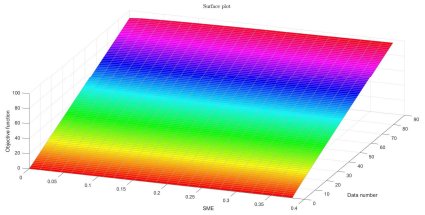


(D) Immunotherapy

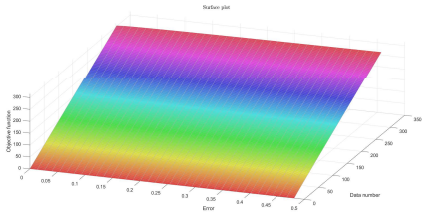
FIGURE 3. Variability in Objective Function



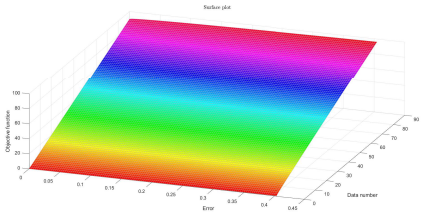
(A) Breast cancer



(B) Cryotherapy



(C) Heart-Cleveland



(D) Immunotherapy

FIGURE 4. Relationship Between Objectives and Input Data

In Figure 4, a clear representation is provided, showcasing how a response variable, specifically the proposed objective function f_2 varies concerning two independent variables. The plot visually communicates the dynamic changes of f_2 across the input space, emphasizing its proximity to the target output.

7. CONCLUSIONS

In this article, a framework named GradiLearn with NSGA-II (GRL-NSGA II) is introduced to achieve an ensemble of classifiers, aiming to enhance accuracy and minimize errors. To achieve the minimization of Mean Squared Error (MSE), a second objective function is proposed. GradiLearn is a combination of a self-adaptive learning rate with conventional Back-propagation Neural Network (BPN). The GRL-NSGA II framework is built upon the ensemble of the GradiLearn with NSGA II, where the initial population is formed using the proposed GradiLearn approach. The framework incorporates a non-cooperative fitness function, Accurate Classification Rate (ACR), to improve the accuracy of classification rates for each class. This approach is specifically applied to address imbalanced pattern classification problems. The

paper also introduces the concept of Canberra distance in crowding distance, namely the CB-Dist based Crowding Distance Algorithm. Experimental results demonstrate that the proposed GRL-NSGA II method achieves superior classification accuracy and other statistical measures compared to existing ensemble methods and conventional machine learning models.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] R. de Albuquerque Teixeira, A.P. Braga, R.H.C. Takahashi, R.R. Saldanha, Improving generalization of MLPs with multi-objective optimization, Neurocomputing. 35 (2000), 189–194. https://doi.org/10.1016/s0925-2312(00)00327-1.

[2] M.I. Al-Janabi, M.H. Qutqut, M. Hijjawi, Machine learning classification techniques for heart disease prediction: a review, Int. J. Eng. Technol. 7 (2000), 5373–5379.

[3] P.A. Castillo, J.J. Merelo, A. Prieto, V. Rivas, G. Romero, G-Prop: Global optimization of multilayer perceptrons using GAs, Neurocomputing. 35 (2000), 149–163. https://doi.org/10.1016/s0925-2312(00)00302-7.

[4] Y. Chen, X. Zhang, Y. Zhuang, B. Yao, B. Lin, Granular neural networks with a reference frame, Knowledge-Based Syst. 260 (2023), 110147. https://doi.org/10.1016/j.knosys.2022.110147.

[5] K. Deb, Multi-objective evolutionary algorithms, in: J. Kacprzyk, W. Pedrycz (Eds.), Springer Handbook of Computational Intelligence, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015: pp. 995–1015. https://doi.org/10.1007/978-3-662-43505-2_49.

[6] K. Deb, Multi-objective optimization, in: E.K. Burke, G. Kendall (Eds.), Search Methodologies, Springer US, Boston, MA, 2014: pp. 403–449. https://doi.org/10.1007/978-1-4614-6940-7_15.

[7] D.E. Goldberg, Genetic algorithms, Pearson education India, (2013).

[8] S. Haykin, Neural networks: a comprehensive foundation, Prentice Hall PTR, (1988).

[9] A.M. Hemeida, S.A. Hassan, A.A.A. Mohamed, et al. Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research, Ain Shams Eng. J. 11 (2020), 659–675. https://doi.org/10.1016/j.asej.2020.01.007.

[10] Y. Jin, Multi-objective machine learning, Springer, Berlin, (2006).

[11] A. Junio Guimarães, P. Vitor de Campos Souza, V. Jonathan Silva Araújo, et al. Pruning fuzzy neural network applied to the construction of expert systems to aid in the diagnosis of the treatment of cryotherapy and immunotherapy, Big Data Cogn. Comput. 3 (2019), 22. https://doi.org/10.3390/bdcc3020022.

[12] S. Khatri, D. Arora, A. Kumar, Implicating machine learning towards development of intelligent system for wart treatment therapy identification, Int. J. Eng. Technol. 7 (2018), 754–758.

[13] R. Abu Khurma, D. Albashish, M. Braik, A. Alzaqebah, A. Qasem, O. Adwan, An augmented Snake Optimizer for diseases and COVID-19 diagnosis, Biomedical Signal Processing and Control 84 (2023) 104718. https://doi.org/10.1016/j.bspc.2023.104718.

[14] A. Layeb, Novel Feature Selection Algorithms Based on Crowding Distance and Pearson Correlation Coefficient, Int. J. Intell. Syst. Appl. 15 (2023), 37–42. https://doi.org/10.5815/ijisa.2023.02.04.

[15] G.P. Liu, Learning with multi-objective criteria, in: 4th International Conference on Artificial Neural Networks, IEE, Cambridge, UK, 1995: pp. 53–58. https://doi.org/10.1049/cp:19950528.

[16] S. Malakar, S.D. Roy, S. Das, et al. Computer based diagnosis of some chronic diseases: a medical journey of the last two decades, Arch. Comput. Methods Eng. 29 (2022), 5525–5567. https://doi.org/10.1007/s11831-022-09776-x.

[17] M.M. Rahman, S. Wang, Y. Zhou, et al. Predicting the performance of cryotherapy for wart treatment using machine learning algorithms, in: Proceedings of the 2019 IISE Annual Conference, May 18-21, 2019, Orlando, FL, USA.

[18] P. Ramdass, G. Ganesan, Leveraging neighbourhood component analysis for optimizing multilayer feedforward neural networks in heart disease prediction, Math. Model. Eng. Probl. 10 (2023), 1317–1323. https://doi.org/10.18280/mmep.100425.

[19] S. Mane, S.S. Sonawani, S. Sakhare, Classification problem solving using multi-objective optimization approach and local search, in: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), IEEE, Chennai, India, 2016: pp. 243–247. https://doi.org/10.1109/ICEEOT.2016.7755105.

[20] M.M. Mijwil, I.E. Salem, R.A. Abttan, Utilisation of machine learning techniques in testing and training of different medical datasets, Asian J. Comput. Inform. Syst. 9 (2021), 29–34. https://doi.org/10.24203/ajcis.v9i4.6765.

[21] D.J. Montana, L. Davis, Training feedforward neural networks using genetic algorithms, in: Proceedings of the 11th International Joint Conference on Artificial Intelligence, Vol. 1, 1989, pp. 762–767.

[22] V. Ravi, D. Pradeepkumar, K. Deb, Financial time series prediction using hybrids of chaos theory, multilayer perceptron and multi-objective evolutionary algorithms, Swarm Evol. Comput. 36 (2017), 136–149. https://doi.org/10.1016/j.swevo.2017.05.003.

[23] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature. 323 (1986), 533–536. https://doi.org/10.1038/323533a0.

[24] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, MIT Press, Cambridge, 1986: pp. 318–362.

[25] B. Sai Lakshmi, G. Gajendran, Evaluation of efficiency of MAMFFN for predicting Caesarean deliveries, AIP Conf. Proc. 2516 (2022), 250016. https://doi.org/10.1063/5.0110051.

[26] B.S. Lakshmi, G. Gajendran, P–dist based regularized twin support vector machine on imbalanced binary dataset, Math. Stat. 11 (2023), 325–334. https://doi.org/10.13189/ms.2023.110211.

[27] K. Senhaji, H. Ramchoun, M. Ettaouil, Training feedforward neural network via multiobjective optimization model using non-smooth $L_{1/2}$ regularization, Neurocomputing. 410 (2020), 1–11. https://doi.org/10.1016/j.neucom.2020.05.066.

[28] K.M.M. Uddin, N. Biswas, S.T. Rikta, et al. Machine learning-based diagnosis of breast cancer utilizing feature optimization technique, Computer Methods Progr. Biomed. Update. 3 (2023), 100098. https://doi.org/10.1016/j.cmpbup.2023.100098.

[29] C.C. Yu, B.D. Liu, A backpropagation algorithm with adaptive learning rate and momentum coefficient, in: Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290), IEEE, Honolulu, HI, USA, 2002: pp. 1218–1223. https://doi.org/10.1109/IJCNN.2002.1007668.

[30] M.H. Zangooei, J. Habibi, R. Alizadehsani, Disease diagnosis with a hybrid method SVR using NSGA-II, Neurocomputing. 136 (2014), 14–29. https://doi.org/10.1016/j.neucom.2014.01.042.