Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Review Optimized Artificial Neural Network by Meta-Heuristic Algorithm and its Applications

Noor Hassan Kadhim[1] , Dr. Qusay Mosa[2]

[1]Computer Science Department, College of Computer Science and Information Technology, University of Al-Qadisiyah , Iraq, E-mail:com.post06@qu.edu.iq

[2]Computer Science Department, College of Computer Science and Information Technology, University of Al-Qadisiyah , , Iraq, E-mail: qusay.mosa@qu.edu.iq

ARTICLE INFO

Article history:
Received: 01 /06/2021
Rrevised form: 12 /06/2021
Accepted : 30 /06/2021
Available online: 07 /08/2021

Keywords:

Artificial neural network; metaheuristic optimizing algorithms; prediction; classification.

ABSTRACT

A Meta-Heuristic Algorithms Optimization (MAHO) is inspired by nature. The Artificial neural network (ANN) has been shown to be successful in a variety of applications, including machine learning. ANNs were optimized using meta-optimization methods to enhance classification performance and predictions. The fundamental objective of combining a meta-heuristic algorithm (MHAO) with an artificial neural network (ANN) is to train the network to update the weights. The training would be speedier than with a standard ANN since it will use a meta-heuristic method with global optimal searching capability to avoid local minimum and will also optimize difficult problems. will discuss some of these meta-heuristic algorithms using ANN as they are applied to common data sets, as well as real-time specific classification and prediction experiences. In order to give researchers motivational insights into their own fields of application.

MSC. 41A25; 41A35; 41A36

DOI : https://doi.org/10.29304/jqcm.2021.13.3.825

1. Introduction

In a variety of fields, new apps are being created to cover daily tasks. As a result, there is a higher need for problem-solving approaches. There have been several machine learning algorithms [1] standards produced during the last few decades.

Prediction Pattern recognition, decision making, and a variety of other applications benefit greatly from these techniques. Regardless of how they are implemented, there is a gap between domain-specific applications and

∗Corresponding author; *Noor Hassan Kadhim*

Email addresses: com.post06@qu.edu.iq

Communicated by*: Dr. Rana Jumaa Surayh aljanabi.*

solution algorithms. To compensate for this shortcoming, these methods have been updated. This results in faster convergence with fewer iterations, increasing the algorithm's efficiency.

Non-linear programming, linear programming, and other traditional optimization techniques, which have been demonstrated to be successful but also have a number of downsides, are nearly unheard of. The deep or baffling power of an objective function or function that fulfills a given constraint in order to establish a suitable aim or value and operate the system effectively is known as optimization. Various Nature Inspired Optimization Algorithms have been created in recent years and effectively utilized to enhance machine learning algorithms [1,2] such as Artificial neural network, Support Machine, Extreme Learning Machine, Deep Learning Machine, and so on. determine the best interpretation based on previously recorded solutions, yet some people have no recollection. Swarm intelligence [3], evolution, physical phenomena, and other MHOAs [4] are the most common. MHOAs are essentially separated into responsibility-based and population-based algorithms [4] depending on the scanning procedure. Touching a candidate solution (a search agent or an item) initiates the decision-making process for a single solution based on an algorithm, which improves over a certain number of repetitions. In any case, a search process for a population-based solution begins with a collection of candidate solutions that are optimized over the required number of iterations, and then the best solution is selected. It was debated in the poll's comment section. [5,6] This survey focuses mostly on population algorithms. Exploration and exploitation are the two key components of MHOA. Diversification is the process of development.

Investigates [7] the whole search area in quest of new, more diversified solutions. The latter is a condensing technique that makes use of the data in the best solution identified throughout the local search region. Selection, crossover, and mutation are evolutionary operations that impact an exploration and exploration method to offer high-quality results. It must be processed via iterations under specific conditions to arrive at the best answer.

MHOAs are often used in conjunction with ANN to pick features for the purpose of determining optimal training parameters and solving other issues that are utilized to determine the best ANN training solution and also its applications to classification and prediction issues. In some cases, even the most successful algorithms might produce poor results in specific applications. This might be due to a mismatch in mixed workbooks for a particular application. To prevent such issues, one needs to focus on selecting a suitable ANN optimization technique and architecture that would function well with a given application in order to have the best results. apply with standard standardized dataset for categorization and prediction gathered from public repositories like the Irvine Machine Repository at the University of California, and some other real-time applications.

A remainder of the paper is laid out as follows: Section 2 shows how to train artificial neural networks using classic optimization methods, and Section 3 discusses how to use meta-heuristic algorithms with artificial neural networks. Section 4 describes generic ANN training using MHOAs, Section 5 explores applications of improved ANN classifiers for stander datasets, and Section 6 addresses single-objective optimization utilizing ANN for specialized real-world problems. And conclusion in Section 7.

An artificial neural network (ANN) is a feed forward neural network (FFNN) design with three layers: input, hidden layer, and output layer. The training of an artificial neural network is a continuous optimization process that involves traveling from input to output to find the best collection of biases and weights in the shortest amount of time. The goal is to improve classification accuracy while lowering classification errors. The performance ANN is solely determined by the synaptic weights. The backpropagation networks [3,4] are a gradient-based learning approach that is commonly used to train artificial neural networks. When using a backpropagation network for training, the input signals are sent in the forward direction and the errors are transmitted in the reverse direction without creating a loop in the network. The errors are calculated as the difference between the output of the target and the output of the network. The backpropagation network training traps the algorithm in a local minimum, resulting in poor convergence rates. In many circumstances [8,9], the MHOA is used with BPN to overcome these restrictions. Figure 1 depicts the construction of ANN [10].

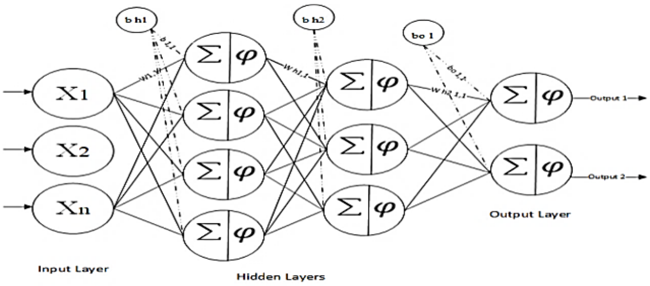

Figure 1: ANN's Architecture

3. Algorithms combining meta-heuristics with artificial neural networks

There are two types of NN learning: supervised learning and unsupervised learning. The most prevalent moderated learning approach is inverse diffusion back-propagation (BP) [11], which is based on the gradient descen (GD) [12] approach and has limitations such as capturing local minimums and sluggish convergence velocity. The major purpose of employing the metaheuristic algorithm instead of the (GD) approach is to train a neural network for weight training updates. The training will be speedier than the classic GD technique, which employs the metaheuristic algorithm's optimal global search capability, avoids local minimums, and improves challenging situations [12].

4. MHOAs are used in a general training method.

The initialization of an ANN with random weights and biases determined by the MHOA is the first stage in ANN training. Training input samples are provided to the network from of the classifications or predictions dataset. The output of an ANN is compared to a goal output. This error value can be calculated using error functions such as mean square error (MSE), sum of square error (SSE), root mean square error (RMSE), and others. In order to achieve the smallest trap value, goal functionality is limited. MHOA is applied on random data in order to generate the next set of weights and distortions for the next iteration. The target function uses input variables to access each potential population. This procedure is repeated until all of the halting requirements have been satisfied. Finally, the ideal solution is the one with the best fit value (lowest error) over all iterations. This is used to classify or forecast data that will not be seen in the near future. A general mass diagram of a classifications or predictions process employing an optimized ANN is shown in Figure 2.

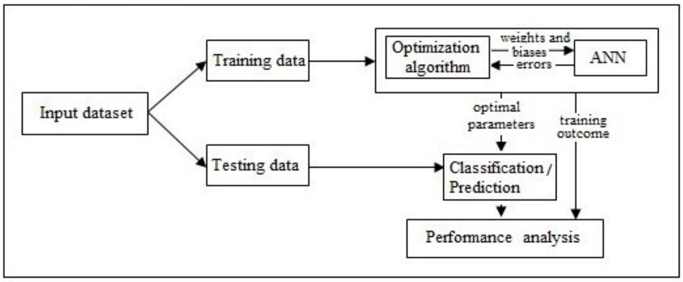

Figure 2: block diagram utilizing optimal ANN in general

5. Some of metaheuristic based on ANN and apply a benchmark dataset

To demonstrate their efficacy, the hybrid classifier was created and used to benchmark classification then prediction records. This section discusses some of these usage in the literature, as well as a quick introduction to the MHOAs employed.

A. optimization of Ant colony with ANN

The search agents in [13] the Ant Colony Optimization (ACO) are called ants, and they replicate the foraging activity of actual ants. This is classified as swarm intelligence. In this procedure, an ant colony is kept and a population is taken into account. Each ant looks for food in the vicinity of its nest at random. It assesses the quantity and quality of food available. When it returns to its nest with food, it will leave pheromones based on this assessment. In order to locate the quickest path between its colony and food supply, this pheromone channel allows for indirect communication with other ants. The combined value supplied to each path is used to assess goal's functionality. After several iterations with newly established ant colonies, the ideal value is discovered. Percentage squared error is a different objective function that is employed. The ACO method [14] is used to choose subsets of features [15], which may then be used to eliminate noise-generating input characteristics. ACO was integrated with traditional gradient-based training methods to create a continuous improvement variation. With [16] sets of sample data as network input for this prediction process, the ACO method was utilized to train ANN to estimate residential building performances depending on the application performances, safety performance, durability, corporate environmental, and economic growth.

B. optimization of Artificial Bee-Colony with ANN

The artificial bee colony (ABC) algorithm [17,18] is inspired by a genuine bee colony's clever search activity. The inhabitants of ABC are made up of three kind of artificial bees: busy bees, who make up half of the population, communicate information about their food source on the dance floor through an oscillating dance, watched bees, who wait in the hive for information about the food source of busy bees, and Scout leader, who is looking for a new food source. When a busy bee's food supply is depleted, this becomes a Scout Bee. The robotic bees' selection of a high-quality food supply is an ideal net solution.

ABC is used to train ANN for a variety of applications, and we'll go through a few of them. [19], [20] In the creation of synaptic weights, structure, and transport function for each neuron, bee colonies are utilized to minimize the number of connections. To test the network, MSE and classification error rate (CER) are utilized. There is also a dimensional decrease. The FFNN is trained using the ABC-based BPN technique [21], and its performance is evaluated using SSE, affinity velocity, and the stability of the optimal solution. For S-system models [22], the ABC [17] technique is utilized to train ANN (ABC-ANN). The biochemical profiles provide details on the network architecture and system dynamics It also aids in the gathering of data for purposes such as determining disease state trends, medication toxicity, and so on. Nonlinear dynamic system that help in the modeling of complicated biological networks is known as systems. ABC-ANN has been shown to be effective in approximating gene expressions.

C. optimization of Bacterial foraging algorithm with ANN

The BFOA algorithm represents the chemical activity of the bacteria in our gut, Escherichia coli. The population is made up of bacteria, and each bacterium serves as a study instrument. Chemical affinity, aggregation, reproduction, elimination, and elimination are all processes in the foraging process that are determined by the control system. The best solution is a bacterial organ that produces the least amount of error during training. [23] BFOA is applicable to FFNN and FF sequencing-generated trains. ANNs are abbreviations for Announcement Numbers. The position of the bacterium with the best fitness is viewed as a collection of ideal values that may be described in terms of precision.

D. optimization of Bat algorithm with ANN

The Bat algorithm (BA) is inspired by the behavior of microorganisms' echolocation after spacing. The population is made up of bat colonies, and each bat is a prospective solution. For ANN training, the BI method is paired with the backpropagation method. [24] To indicate a solution, place a food supply in that spot. To create a better balance between exploration and exploitation, several cooperative techniques, such as master and slave strategies collaborating with the paddle algorithm, are used to increase collaboration amongst population groups. The Bat approach is then used to optimize the ANN design in terms of feature selection, hidden layer number, hidden layer node number, and hidden layer weight. In place of the [25] random numbers, the random card was employed. The Bat method is commonly utilized in medical image analysis to tackle real-world global optimization problems.

E. Optimization of Biogeography algorithm with ANN

BBO stands for biogeography-based optimization, which is an evolutionary method based on biological species' geographic distribution. Populations for the algorithm may be found in a variety of environments. Habitat is research instruments that evolve as a result of migration and mutations. The Habitats Suitability Index are used to determine whether habitats are suitable for study (HSI). Residents in high-HSI habitats move to low-HSI habitats, resulting in a high rate of migration in low-HSI environments. The BBO algorithm optimizes HSI for all environments.

Each environment is an MLP with its own rates of movement, immigration, and reciprocity. The type of difficulty determines the number of habitat (weight and distortion) in the habitats. The MSE of all training techniques for all habitats is used to calculate the HSI, which is a fitness function. As a result, habitats are grouped together based on migration and migration rates. This suggests that the environment varies in response to the rate of mutation. The greatest way to conserve habitat for the future generation is to have a solid HSI [26].

The BBO method is used to the application of fruit classification and is utilized to optimize FFNN with training parameters [27]. 1653 color fruit photos from 18 classes were used as input for categorization. The BBO-FNN classifier is used to preprocess images in four phases, extract features, assess significant components, and ultimately classify them. BBO-FNN has been seen to outperform the others. The next use of BBO-trained MLP will be to detect spam emails. During training, the BBO-MLP [28] sample is used as a filter to extract information about spam emails. In comparison to previous classifiers, the model distinguishes spam emails from ham emails with the maximum accuracy using the knowledge gathered during training.

F. optimization Bird-mating with ANN

Bird Mating Optimizing (BMO) is a program inspired by bird mating behavior. Monogamous, polygamous, polyandry, and non-mixed birds form a society, which is made up of four types of birds. Each bird is a genetic experiment with a certain number of genes. During the breeding season, birds participate in the mating process in order to breed with superior genes. The objective function or fitness function that governs the mating strategy is used to determine the quality of each brood. The birds are sorted and placed into four classes based on their quality. The pairing procedure is then carried out. The brood's quality is compared to that of the bird, and if the brood is superior, the bird is replaced by the brooder. BMO is a tool for locating weights for exercise [29]. ANN. Every bird is a potential solution to the issue. The number of genes in a bird is determined by the problem's variations and defines the dimension d wherein reproduction occurs. The quality of the bird is determined by its gene, and if it is higher than that of its brooder, the brood is expelled from the society. If this is not the case, the reverse is done. For decades to come, this will be repeated. Finally, for ANN training, a strong genetic quality brood or bird is the best weight solution.

G. optimization of Firefly with ANN

The process of optimization to attract mating partners and prospective prey, the Firefly Algorithm (FA) [30,31] uses the scintillation features of fireflies created by the bioluminescence process. The population is made up of a swarm of fireflies, with each fly acting as a search agent. As the distance between the fireflies grows, the light intensity drops The lens function is assigned to the blinking light. The following are the three ideal guidelines for characterizing FIFA: (1) Because they are thought to be of two sexes, all fireflies are attracted to one another. (2) The movement of fireflies in dimensional space is proportional to their brightness, and (3) the brightness of firefly is controlled by the landscape of the objective function. To determine the gravity, the new solution is reviewed and the light intensity is updated each time. The finest option is Firefly, which delivers the greatest options. To boost its performance index, FA is employed in combination with BPN. The performance indicator is calculated using SSE. The most appealing firefly is deemed optimal since it has a high light intensity. Two transmission functions are employed to assess the FA, which is utilized to train FFNN. [32] As an error function, MSE is utilized. The FA is assessed depending on the error rate and is utilized in conjunction with the artificial neural network to solve the time series classifications issue. ANN models such as the Link ANN function and the Radial Base Function Network use the Firefly [33] algorithm. This hybrid method forecasts software expenses regardless of their complexity. When employing data records from the PROMISE software repository, accuracy has been demonstrated to improve.

H. Genetic algorithm with ANN

Based on Charles Darwin's "Theory of Natural Selection and Genetics," John Henry Holland and his team developed a gene algorithm. [34, 35] (GA). To address the challenge, simulation relies on the strongest members of the population surviving for the largest number of generations. The population is made up of required number of persons or chromosomes. Every chromosome is a possible solution, and it is made up of a particular amount of genes that are related to a degree of fitness depending on the problem. The algorithm evolves through three main processes that are influenced by reproductive techniques including selection, crossover, and mutation. Parents are picked for their fitness values when it comes to mating. The best chromosome or child with excellent fitness will be utilized as an ideal solution to address the problem after a set number of generations. ANN has already been used in combination based GA (GA-ANN) in a number of articles. For classification jobs, GA-ANN has been used in some cases. [36] GA is used to train ANN on the best chromosome with a high fitness score, and the method's effectiveness is proved by comparing its results to MLP trained using the classic BPN approach. [37] classification error ratio (CEP) and The squared error ratio (SEP) are utilized to compare the algorithm's efficiency to that of an ANN built with BPN and trained using a traditional LM.

When a light hydrocarbon escapes and is retrieved, a free product, such as oil, will scatter into the water surface. This technique needs a comprehensive set of input parameters. Inverse modeling and other parameter estimate approaches can help with this. To tackle this problem, enhanced ANN was combined with GA (ANNGA).

I. optimization Grey-wolf with ANN

Gray Wolf Optimizing (GWO) [38] is a gray wolf commanding and hunt system hierarchically. Alpha, Beta, Delta, and Omega are four classifications of gray wolves (searching agents) used in the initial simulation. The fitness values of wolves are examined using the objective function once they have formed a population. As a result, the first three wolves are classified as Alpha, Beta, and Delta in the system, with high physical fitness, while the rest of the wolf pack is classified as an omega, that is change data on first three wolves to attack a victim's sites. Finding, circling, and attacking prey are the three main steps in prey hunting. Placement Alpha is, after all, the most effective optimization method. MLP is trained with optimum parameters using the GWO. The average calculation is [39] target locations.

J. optimization algorithm of Krill Herd with ANN

Krill Herd algorithm (KHA) [40] is depended on krill herd behaviors. The entire krill flock is referred to as a population. The ideal option is to keep the krill close to the best food supply and to use a high-intensity burner with a low goal function. Three factors, such as induced motility, foraging motility, somatic motility, and prevalence, all contribute to this movement. To boost its performance, hybridization and mutation agents are used. After each iteration, the location of each krill is updated. The better closing positioning of the krill after maximum iterations is the optimal solution, the global ideal solution works on a basis of as a set of parameters Training ANN. The trapping function [41,42] is MSE51minimum and SSE52 minimum.

K. optimization algorithm of Moth-flame with ANN

The moth flame optimizing (MFO) [43] technique is depended on the butterfly's navigation ability, which is used to generate the population. Every month, a search agent will be assigned. Through feature known as transverse orientation, the moth fly towards the moon (flame) at angle in a straight line over extended distances at night. Because the distance between them and the moon is so short, they move in a spiral pattern towards artificial light. The moths, who make up the total population, are the search agents in this algorithm. Candidate solutions are the moths, and their placements define fitness values. In each cycle, moth locations are updated as better solutions are discovered based on the flame with a higher fitness rating. In terms of termination, the moth in the best position is seen to be the best option. The MFO technique is used to train the MLP with the best moth. of the minimum average MSE is the objective function used to find the best solution [44].

L. optimization of Particle-swarm with ANN

Particle swarm optimizing (PSO) [45] [10] is dependent on bird flying behavior. The search agents are the algorithm community's flocks of birds (also known as particles). To avoid collisions, a flock of birds that feed on birds maintains a safe distance between them while flying. These birds in the group share information about where they may get food. The best value (pbest) and location are remembered by each bird (pbestx and pbesty). *They adjust their X (set to adverse random weight when motion is to the right of Pbestx; alternatively, increase a random weight if motion is to the left) and Y (up and down to move up and down) speeds based on their position throughout sudden movements and changing direction to keep the herd moving. Each bird also recalls the best international location for a bird that found the best value (gbest). Finally, after the maximum number of iterations, the flock of birds arrives at the optimal location of the vortex's food supply. Flying with Gbest is thought to be the most efficient way to progress. The optimum parameters acquired from the PSO are also used to train the ANN.*

To train ANN to forecast, an opposition-dependent particle swarm optimization approach is utilized in combination with BPN. [46] Opposition-Based Learning is a framework for learning to gaze in the opposing way. In the opposed particle swarm optimization process, the particle with the lowest fitness value is replaced by its opposing particles. The CAPSO (Accelerated Central [47] Particle Gravity) optimization technique, which was created to merge Newton's laws of motion based PSO, is an upgraded PSO approach. This enhances the algorithm's learning and convergences speed. Additional parameters depended on Newton's principles of motion in mechanical have been incorporated to the software. The CASPO is used in conjunction with the ANN to identify medical conditions.

The PSO algorithm is updated as a self-adaptive chaotic PSO method (CSPSO)To improve performance, speed convergence, and avoid early convergence. CSPSO uses chaotic sequences based on chaos theory to introduce an adaptive inertia weight factor and modification of acceleration coefficients. CSPSO is combined with the BP algorithm to train artificial neural networks and predict gas solubility in polymers at various pressures and temperatures. The accuracy and predictability of this predictive model have been demonstrated.

M. Wolf search algorithms with ANN

The Wolf Search Algorithm (WSA) [48] is depended on wolves' predatory behaviors and their predatory defense systems. The population is built up by wolf herds, and every wolf is search agent. The algorithm employs three criteria when determining the optimal solution.

(1) Every wolf is aware of it is companion and plan from the outset.

(2) The target function was used to determine a quality of the wolf's present position. The decision of the wolf to move is calculated.

(3) The wolf is no longer visible.

ANN training results in BPN weight. [49] searching for the best solution in a number of methods Weights are computed, validated, and compared to the best answers in backdrop renderings. It was repeated until you reach the MSE or STOP state.

N. Dragonfly algorithm with ANN

The Dragonfly [50] (DA) algorithm was created using two different types of Dragonfly swarm behavioral patterns: dynamic and static. A swarm of dragonflies is a group of organisms in which each dragonfly acts as search agent. Dragonfly is drawn to food source and serve as a deterrent to adversaries. Separation, alignment, cohesion, feeding, and galloping are five important criteria that the dragonfly uses to update its location based on these actions. The first two elements are dealt with by a static swarm, while the remaining agents are dealt with by a dynamic swarm. By changing these elements for improvement, the trade-off between exploration and exploitation is maintained. The optimal answer is provided by the ideal location of a dragonfly. The fundamental fuel requirements in India are predicted using a trained ANN with the optimal global parameter (low MSE) determined from DA.

O. optimization of Monarch butterfly with Artificial neural network

The Monarch Butterfly Organization [51] (MBO) is based on monarch butterfly migratory behavior from one nation to another at specific periods of the year. The population is made up of a flock of monarch butterflies, each of which serves as a search agent. There are two subgroups in the whole population. The two operators utilized in algorithm to modify a position of monarch butterfly are the relays factor and monarch adjustment trigger. This process is repeated until the optimal option is discovered. To identify osteoporosis, the MBO algorithm is used with ANN [52].

Table 1: Different Optimizers are compared.

Algorithm	Advantages	Disadvantages
ACO	Helpful in improving weight	Indefinite times for convergence
GA	It improves computing speed through parallel search	The exact solutions can't always be found and it complicated Algorithms
BBO	Has a balanced capacity to handle high-dimensional problems while avoiding local minima.	Bad solution
BAT	is a very promising algorithm because its implementation and comparison	It consumes a lot of memory
PSO	It is also easy to calculate and implement	Poor local search abilities
ABC	It commands great attention for its rapid problem solving, greedy research guideline, and strong global convergence	search space is limited by initial solution

6-Optimizations with a single goal (SO):

Often real-world issues are defined in order to find the best solutions for achieving the best values of single objective function which combines all of goals into one. Goal orientations optimizing is a strong technique that provides insight into the characteristics of an issue to the decision maker.

Table 2 some of MHAO single objective with Artificial neural network

Ref.	Methods	Dataset	instance	Attribute	classes	Objective	NN-models	Constraint
[53]	Bat-DNN, MBat-DNN, Mean Bat-DNN (Piecewise-Bat-DNN, Logistic Bat-DNN, Sin Bat-DNN)	IRIS Diabetic Thyroids Cancers Cards Glasses Mackey-Glasses GAS Furnaces	150 768 7200 699 690 214 1000 296	5 7 22 11 14 11 2 3	3 2 3 2 2 6 0 0	"MSE ¼ N1 PNffi1^{N}i¼P1Niyi1 byi i^{2}b, i2 RMSE ¼ q ¼ y y , and The model's utilization of a percentage of the total number of connections."	Hyperbolic tangent are used as activation functions.	"30 Two-fold iteration, Size of population = 100, Max numbers of iteration = 100, Min frequency = 0, Max frequency = 2, Loudness of emission = 0.25, Pulse rate = 0.5, Probability = 0.5"
[54]	HPABC... hybrid PSO and ABC algorithms	IRIS Cancers Diabetic Glasses	150 699 768 214	5 10 9 8	3 2 2 6	"SSE ¼ 1n Pni¼1ðti yi Þ 2, n is number of patterns in the specified data set, t is target of first pattern, y is actual result of average elapsed time for the first pattern, the average job score, means is classification accuracy"	"3 Layers NN, iris (4-5-3), cancer (9-8-2), diabetes (8-7-2), and glasses (9-126), sigmoid activation function "	"Compared based original ABC & PSO, Unlike ABC algorithm, HPABC, GA, BP, MWO, HS-BTW and EMWO have 6 tuned parameter SN and w, learning factor, value of limit, maximum number of cycles and 30 iterations, SSE = 0.0001"
[55]	BAT-NN algorithm	Glasses Diabetic Cancers Thyroids	214 699 768 7200	8 8 6 20	6 2 2 3	"MSE ¼ $^{1}{}_{n}$ PnI 1$^{n}{}_{i}$¼1ðAi YiÞ^{2}, CEP ¼ Pn¼ incorrect instances=n, ACC ^{¼} PI¼1correctinstances=n"	MLP-NN	"No. of iterations = 30, no. of bats = 10, loudness varies from a large positive A_0 to a minimum constant A_{min}, r = 0.5, A_0 ¼ 0:5, f ¼ ½0; 2icomparedto ABC, PSO, GA, and BP learning algorithm, and 10 fold cross validation"
[56]	ACO-BP algorithm	WBC Diabetic Thyroids Hearts Genes Soybeans Cards Glasses Horses	699 768 920 7200 3175 364 690 214 683	9 8 35 21 120 58 51 9 35	3 3 1 2 3 4 5 20	"Percentage of classifications error CEP 100 for test pattern with incorrectly classified pattern, squared error E ¼ Pnp¼1Poi¼1tpi opi 2"	"FFNN 9-6-2 for WBC, 8-6-2 for Diabetes, 35-6-2 for Heart, 21-6-3 for Thyroid, 120-6-3 for Gene, 58-6-3 for Horse, 51-6-2 for Card, 9-6-6 for Glass, 82-6-19 for Soybean"	"50 runs, 1000 function evaluation, compared BP, LM, GA, RCH, ACO_R, PSO, and ABC algorithm, m = 2, a probability (p) 1 –q, where q (0 q 1) is parameter of the decision rule, ^{q} ¼ 0:05, g = 0.01, b (the factor for LM = 50 in Cancer, 5 in Diabetes, and 1.5 in Heart), d = 30, 4-fold cross validation"

UCI Lab [57] was used to maintain four standardized classification records in an experimental setting. A comparison was done between the suggested hybrid algorithm and the original ABC and PSO algorithms to demonstrate the resilient performance of hybrid technologies. HPABC is a hybrid algorithm that combines the benefits of ABC and PSO algorithms and integrates them. Randomly configured solutions were created in the previous phase of HPABC. Each solution was represented by the dimension vector D xid, with I = {1... SN}, SN denoting the population solution number, and d denoting the variable number to be improved. Table 3 shows the topology of the neural network for each standard data set. Because the cancer dataset has a suitable degree of complexity in terms of ANN structure and a significant number of instances compared to the other datasets, the optimum factor settings derived from it were generalized. We may infer that while the HPABC approach struggled with a large number of classes and limited datasets (Glass), it could be utilized to train ANN well.

presented an enhanced bat methodology with a single aim and some modified solution representations to tackle classification and prediction problems using ANN structure and weights [53]. The best single solution, best single technique, and best global solution, as well as three chaotically planned applications to construct a chaotic sequence instead of any sequence [58] are the basics of the original bat algorithm and some of the modified variants addressed in this work. When compared to other forms, the methodology variants produced correct results and then were matched to use the Taguchi technique, a stochastic approach proved in dynamical and nonlinear systems. Diversification of each solution was done by altering the position, speed, and frequency of the bat population. In terms of training errors, test errors, and connection, Logistic Bat-DNN placed best in all scenarios, beating other gas oven data sets approaches in predicting time series. Table 3 shows the specifics of this strategy. Al-Nuaimi et al. proposed utilizing FFNN to implement the HPABC (Synthetic Single-Targeted Particle Hybrid Bee Colony) method in [54].

A single-goal Ant colony optimizing (ACO) method then hybrid ACO method were used to training the frontal neural network to categorize patterns by selecting various combinations of compounded weights value in this study [53]. Several typical issues of the dataset were selected to be addressed in tested trials, as indicated in Table 3. ACO was taught to choose the optimal neighborhood, and then a graded descent method was used to find it. The work proposed in [59] was built on in ACO training, and it has been broadened and investigated further. The reverse diffusion method was performed after each iteration of the ACO method, rather than after the ACO algorithm was completed, which was the key difference and extensibility in this work. According to the findings of the experiments, ACO was a suitable choice for selecting excellent BP values. Standalone ACO outperformed standalone ACOR, while hybrid ACO-BP outperformed hybrid ACO-BP, particularly in major problem scenarios. When the issue size was raised, the gradient descent approaches performed poorly compared to the ACO-BP hybrid algorithms in terms of behavioral performance.

The quality of the data has a major impact on the efficacy of Artificial Neural Network (ANN) training, and this quality is determined by specific parameters such as accuracy, adaptability, and precision. The solution to quality issues is data cleaning. Data cleaning is a technique for detecting incorrect or incomplete data and then improving its quality by fixing any problems discovered. Data purification affects ANN training data, thus it should be utilized before ANN modeling ANN learning is also influenced by the data, which may be incorrect or huge. However, because ANN is used to tackle a variety of real-world issues, researchers are experimenting with a variety of ways to enhance it. By designing its techniques, metaheuristics primarily deal with difficulties of generalization and local minimum. The exploration capacity of traditional ANN training approaches such as gradient descent and backpropagation is lacking, resulting in local minimum novel solutions to optimization issues. Meta-heuristics algorithms are superior to traditional methods in terms of exploitation and exploration, and they can be applied to any aspect of ANN. For ANN training, meta-heuristics have been effectively used. These nature-inspired algorithms can be used to identify the global best solution. The ant colony optimization (ACO) method is useful for optimizing weights, whereas the genetic algorithm (GA) method improves computational speed by using a parallel search Biogeography-based optimizing (BBO) excels at solving problems with several dimensions while avoiding local minima. Because of its execution and comparison, the Bat algorithm (BAT) is very promising. The bat algorithm's convergence rate can be affected by the right selection of its parameters. Particle Swarm Optimizations (PSO) is a technique that may be used in both engineering and scientific studies. It also features a basic computation and a quick search feature. The Artificial Bee Colony algorithm (ABC) has received a lot of attention because of its fast issue solving, greedy heuristic search feature, and high global convergence. The variations between different optimizers are shown in Table 2. The results of the experiments showed that each optimization algorithm was a helpful and appropriate heuristic approach for data categorization issues and that it was frequently used for a specific problem category**.** As a result, there is no universal optimizer that can solve all types of problems**.**

7.Conclusion

To produce high-quality solutions, a proper trade-off should be maintained. During the study phase, some MHOAs demonstrate their effectiveness in exploration, while others demonstrate their strength in exploitation. Algorithms employ MHOAs to create ANN optimization approaches in order to maintain a satisfactory result. This paper demonstrates some of the algorithms that have been constructed by combining ANNs with different MHOAs. To benchmark data sets and other studies, surveys are conducted and used in real time. Some of them have been highlighted, which will assist researchers better understand how MHOAs can be employed to improve the ANN. Each algorithm has its own set of benefits and drawbacks since no algorithm can handle all issues. This paper outlines how to select the most appropriate MHOA for a certain application. Several novel MHOAs that have never been used with ANN might be discovered and tested. This approach will offer greater results than the others.

References

[1] H. Alkabbani, A. Ahmadian, Q. Zhu, and A. Elkamel, “Machine Learning and Metaheuristic Methods for Renewable Power Forecasting: A Recent Review,” *Front. Chem. Eng.*, vol. 3, no. April, pp. 1–21, 2021, doi: 10.3389/fceng.2021.665415.

[2] M. Eshtay, H. Faris, and N. Obeid, “Metaheuristic-based extreme learning machines: a review of design formulations and applications,” Int. J. Mach. Learn. Cybern., vol. 10, no. 6, pp. 1543–1561, 2019, doi: 10.1007/s13042-018-0833-6.

[3] K. Kaur and Y. Kumar, "Swarm Intelligence and its applications towards Various Computing: A Systematic Review," 2020 INTERNATIONAL CONFERENCE ON INTELLIGENT ENGINEERING AND MANAGEMENT (ICIEM), 2020, pp. 57-62, doi: 10.1109/ICIEM48762.2020.9160177.

[4] H. Stegherr, M. Heider, and J. Hähner, “Classifying Metaheuristics: Towards a unified multi-level classification system,” *Nat. Comput.*, vol. 0, 2020, doi: 10.1007/s11047-020-09824-0.

[5] K.-L. Du, M. N. S. Swamy, and others, “Search and optimization by metaheuristics,” *Tech. Algorithms Inspired by Nat.*, 2016.

[6] S. Koziel and A. Bekasiewicz, *Multi-objective design of antennas using surrogate models*. World Scientific, 2016.

[7] N. T. Dat, C. V. Kien, H. P. H. Anh and N. N. Son, "Parallel Multi-Population Technique for Meta-Heuristic Algorithms on Multi Core Processor," 2020 5TH INTERNATIONAL CONFERENCE ON GREEN TECHNOLOGY AND SUSTAINABLE DEVELOPMENT (GTSD), 2020, pp. 489-494, doi: 10.1109/GTSD50082.2020.9303114.

[8] T. L. Fine, *Feedforward neural network methodology*. Springer Science & Business Media, 2006.

[9] D. Graupe, *Principles of artificial neural networks*, vol. 7. World Scientific, 2013.

[10] A. H. Alsaeedi, A. H. Aljanabi, M. E. Manna, and A. L. Albukhnefis, “A proactive meta heuristic model for optimizing weights of artificial neural network” 2020, doi: 10.11591/ijeecs. v20.i2. pp976-984.

[11] M. Gori and A. Tesi, “On the problem of local minima in backpropagation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 1, pp. 76–86, 1992.

[12] C. Burges *et al.*, “Learning to rank using gradient descent,” in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 89–96.

[13] M. Dorigo, V. Maniezzo, and A. Colorni, “Ant system: optimization by a colony of cooperating agents,” *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 26, no. 1, pp. 29–41, 1996.

[14] K. Socha and C. Blum, “An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training,” *Neural Comput. Appl.*, vol. 16, no. 3, pp. 235–247, 2007.

[15] R. K. Sivagaminathan and S. Ramakrishnan, “A hybrid approach for feature subset selection using neural networks and ant colony optimization,” *Expert Syst. Appl.*, vol. 33, no. 1, pp. 49–60, 2007.

[16] H. Shi and W. Li, “Artificial neural networks with ant colony optimization for assessing performance of residential buildings,” in *2009 International Conference on Future BioMedical Information Engineering (FBIE)*, 2009, pp. 379–382.

[17] D. Karaboga and B. Basturk, “A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm,” *J. Glob. Optim.*, vol. 39, no. 3, pp. 459–471, 2007.

[18] Taher, A., & Kadhimb, S. (2019). Hybrid between Genetic Algorithm and Artificial Bee Colony for Key Generation Purpose. Journal of Al-Qadisiyah for Computer Science and Mathematics, 11(4), Comp Page 37-46. https://doi.org/10.29304/jqcm.2019.11.4.627

[19] D. Karaboga and C. Ozturk, "Neural networks training by artificial bee colony algorithm on pattern classification," *Neural Netw. World*, vol. 19, no. 3, p. 279, 2009.

[20] B. A. Garro, H. Sossa, and R. A. Vázquez, "Artificial neural network synthesis by means of artificial bee colony (abc) algorithm," in *2011 IEEE Congress of Evolutionary Computation (CEC)*, 2011, pp. 331–338.

[21] S. Nandy, P. P. Sarkar, and A. Das, "Training a feed-forward neural network with artificial bee colony based backpropagation method," *arXiv Prepr. arXiv1209.2548*, 2012.

[22] W.-C. Yeh and T.-J. Hsieh, "Artificial bee colony algorithm-neural networks for S-system models of biochemical networks approximation," *Neural Comput. Appl.*, vol. 21, no. 2, pp. 365–375, 2012.

[23] I. A. A. Al-Hadi, S. Z. M. Hashim, and S. M. H. Shamsuddin, "Bacterial Foraging Optimization Algorithm for neural network learning enhancement," in *2011 11th international conference on hybrid intelligent systems (HIS)*, 2011, pp. 200–205.

[24] N. M. Nawi, M. Z. Rehman, and A. Khan, "A new bat based back-propagation (BAT-BP) algorithm," in *Advances in Systems Science*, Springer, 2014, pp. 395–404.

[25] N. S. Jaddi, S. Abdullah, and A. R. Hamdan, "Multi-population cooperative bat algorithm-based optimization of artificial neural network model," *Inf. Sci. (Ny).*, vol. 294, pp. 628–644, 2015.

[26] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Let a biogeography-based optimizer train your multi-layer perceptron," *Inf. Sci. (Ny).*, vol. 269, pp. 188–209, 2014.

[27] Y. Zhang, P. Phillips, S. Wang, G. Ji, J. Yang, and J. Wu, "Fruit classification by biogeography-based optimization and feedforward neural network," *Expert Syst.*, vol. 33, no. 3, pp. 239–253, 2016.

[28] A. Rodan, H. Faris, J. Alqatawna, and others, "Optimizing feedforward neural networks using biogeography based optimization for e-mail spam identification," *Int. J. Commun. Netw. Syst. Sci.*, vol. 9, no. 01, p. 19, 2016.

[29] A. Askarzadeh and A. Rezazadeh, "A new heuristic optimization algorithm for modeling of proton exchange membrane fuel cell: bird mating optimizer," *Int. J. Energy Res.*, vol. 37, no. 10, pp. 1196–1204, 2013.

[30] A. Askarzadeh and A. Rezazadeh, "Artificial neural network training using a new efficient optimization algorithm," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 1206–1213, 2013.

[31] X.-S. Yang, "Firefly algorithms for multimodal optimization," in *International symposium on stochastic algorithms*, 2009, pp. 169–178.

[32] S. Nandy, P. P. Sarkar, and A. Das, "Analysis of a nature inspired firefly algorithm based back-propagation neural network training," *arXiv Prepr. arXiv1206.5360*, 2012.

[33] I. Brajevic and M. Tuba, "Training feed-forward neural networks using firefly algorithm," *Recent Adv. Knowl. Eng. Syst. Sci.*, 2013.

[34] M. Alweshah, "Firefly algorithm with artificial neural network for time series problems," *Res. J. Appl. Sci. Eng. Technol.*, vol. 7, no. 19, pp. 3978–3982, 2014.

[35] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," *Addion wesley*, vol. 1989, no. 102, p. 36, 1989.

[36] A. J. F. van Rooij, R. P. Johnson, and L. C. Jain, *Neural network training using genetic algorithms*. World Scientific Publishing Co., Inc., 1996.

[37] U. Seiffert, "Multiple layer perceptron training using genetic algorithms," in *ESANN*, 2001, pp. 159–164.

[38] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[39] S. Mirjalili, "How effective is the Grey Wolf optimizer in training multi-layer perceptrons," *Appl. Intell.*, vol. 43, no. 1, pp. 150–161, 2015.

[40] A. H. Gandomi and A. H. Alavi, "Krill herd: a new bio-inspired optimization algorithm," *Commun. nonlinear Sci. Numer. Simul.*, vol. 17, no. 12, pp. 4831–4845, 2012.

[41] N. S. Lari and M. S. Abadeh, "Training artificial neural network by krill-herd algorithm," in *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, 2014, pp. 63–67.

[42] P. A. Kowalski and S. Łukasik, "Training neural networks with krill herd algorithm," *Neural Process. Lett.*, vol. 44, no. 1, pp. 5–17, 2016.

[43] S. Mirjalili, "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm," *Knowledge-based Syst.*, vol. 89, pp. 228–249, 2015.

[44] W. Yamany, M. Fawzy, A. Tharwat, and A. E. Hassanien, "Moth-flame optimization for training multi-layer perceptrons," in *2015 11th International computer engineering Conference (ICENCO)*, 2015, pp. 267–272.

[45] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, 1995, vol. 4, pp. 1942–1948.

[46] M. Yaghini, M. M. Khoshraftar, and M. Fallahi, "A hybrid algorithm for artificial neural network training," *Eng. Appl. Artif. Intell.*, vol. 26, no. 1, pp. 293–301, 2013.

[47] Z. Beheshti and S. M. H. Shamsuddin, "CAPSO: centripetal accelerated particle swarm optimization," *Inf. Sci. (Ny).*, vol.

258, pp. 54–79, 2014.

[48] R. Tang, S. Fong, X.-S. Yang, and S. Deb, "Wolf search algorithm with ephemeral memory," in *Seventh international conference on digital information management (ICDIM 2012)*, 2012, pp. 165–172.

[49] N. M. Nawi, M. Z. Rehman, and A. Khan, "WS-BP: An efficient wolf search based back-propagation algorithm," in *AIP Conference Proceedings*, 2015, vol. 1660, no. 1, p. 50027.

[50] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.

[51] G.-G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 1995–2014, 2019.

[52] D. Devikanniga and R. J. S. Raj, "Classification of osteoporosis by artificial neural network based on monarch butterfly optimisation algorithm," *Healthc. Technol. Lett.*, vol. 5, no. 2, pp. 70–75, 2018

[53] N. S. Jaddi, S. Abdullah, and A. R. Hamdan, "Optimization of neural network model using modified bat-inspired algorithm," *Appl. Soft Comput.*, vol. 37, pp. 71–86, 2015.

[54] L. L. Kamal and H. Kodaz, "Training artificial neural network by bat optimization algorithms," *Int. J. Adv. Comput. Eng. Netw.*, vol. 5, no. 8, pp. 53–56, 2017.

[55] M. Mavrovouniotis and S. Yang, "Training neural networks with ant colony optimization algorithms for pattern classification," *Soft Comput.*, vol. 19, no. 6, pp. 1511–1522, 2015.

[56] A. Asuncion and D. Newman, "UCI machine learning repository." Irvine, CA, USA, 2007.

[57] A. H. Gandomi, X.-S. Yang, A. H. Alavi, and S. Talatahari, "Bat algorithm for constrained optimization tasks," *Neural Comput. Appl.*, vol. 22, no. 6, pp. 1239–1255, 2013.

[58] V. N. Ghate and S. V Dudul, "Optimal MLP neural network classifier for fault detection of three phase induction motor," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3468–3481, 2010.

[59] M. Mavrovouniotis and S. Yang, "Evolving neural networks using ant colony optimization with pheromone trail limits," in *2013 13th UK Workshop on Computational Intelligence (UKCI)*, 2013, pp. 16–23.