Neural Processing Letters
https://doi.org/10.1007/s11063-022-11055-6

Application of Meta-Heuristic Algorithms for Training Neural Networks and Deep Learning Architectures: A Comprehensive Review

Mehrdad Kaveh[1] · Mohammad Saadi Mesgari[1]

Accepted: 11 October 2022

Abstract

The learning process and hyper-parameter optimization of artificial neural networks (ANNs) and deep learning (DL) architectures is considered one of the most challenging machine learning problems. Several past studies have used gradient-based back propagation methods to train DL architectures. However, gradient-based methods have major drawbacks such as stucking at local minimums in multi-objective cost functions, expensive execution time due to calculating gradient information with thousands of iterations and needing the cost functions to be continuous. Since training the ANNs and DLs is an NP-hard optimization problem, their structure and parameters optimization using the meta-heuristic (MH) algorithms has been considerably raised. MH algorithms can accurately formulate the optimal estimation of DL components (such as hyper-parameter, weights, number of layers, number of neurons, learning rate, etc.). This paper provides a comprehensive review of the optimization of ANNs and DLs using MH algorithms. In this paper, we have reviewed the latest developments in the use of MH algorithms in the DL and ANN methods, presented their disadvantages and advantages, and pointed out some research directions to fill the gaps between MHs and DL methods. Moreover, it has been explained that the evolutionary hybrid architecture still has limited applicability in the literature. Also, this paper classifies the latest MH algorithms in the literature to demonstrate their effectiveness in DL and ANN training for various applications. Most researchers tend to extend novel hybrid algorithms by combining MHs to optimize the hyper-parameters of DLs and ANNs. The development of hybrid MHs helps improving algorithms performance and capable of solving complex optimization problems. In general, the optimal performance of the MHs should be able to achieve a suitable trade-off between exploration and exploitation features. Hence, this paper tries to summarize various MH algorithms in terms of the convergence trend, exploration, exploitation, and the ability to avoid local minima. The integration of MH with DLs is expected to accelerate the training process in the coming few years. However, relevant publications in this way are still rare.

✉ Mehrdad Kaveh
m.kaveh11@email.kntu.ac.ir

✉ Mohammad Saadi Mesgari
mesgari@kntu.ac.ir

[1] Department of Geodesy and Geomatics, K. N. Toosi University of Technology, Tehran 19967-15433, Iran

Published online: 31 October 2022

Keywords Deep learning (DL) · Artificial neural networks (ANN) · Meta-heuristics (MH) · Hyper-parameters optimization · Training · And gradient-based back propagation (BP) learning algorithm

Abbreviations

AE	Autoencoder
ABC	Artificial bee colony
ANFIS	Adaptive network fuzzy inference system
ACO	Ant colony optimization
ANN	Artificial neural network
ACS	Artificial cooperative search
BM	Boltzmann machine
AI	Artificial intelligence
BNN	Biological neural network
BA	Bat algorithm
BP	Backpropagation
BBO	Biogeography-based optimization
BRNN	Bayesian regularisation neural network
BMO	Bird mating optimizer
CNN	Convolutional neural network
CCA	Convex combination algorithm
CPNN	Condensed polynomial neural network
CMA-ES	Covariance matrix adaptation based evolutionary strategy
DAE	Deep autoencoder
ChOA	Chimp optimization algorithm
DBM	Deep Boltzmann machine
CRO	Coral reef optimization
DBN	Deep belief network
CS	Cuckoo search
DDAE	Deep denoising autoencoder
DE	Differential evolution
DENNs	Differential equation neural networks
DGO	Dynamic group optimisation
DL	Deep learning
EA	Evolutionary algorithm
DNN	Deep neural networks
EBO	Ecogeography-based optimization
DSN	Deep stacking network
EC	Evolutionary computation
EDEN	Evolutionary deep networks
EvoDL	Evolutionary deep learning
FFNN	Feed forward neural network
EO	Extremal optimization
FLNFN	Functional-link-based neural fuzzy network
ES	Evolution strategy
GAN	Generative adversarial network

FA	Firefly algorithm
GRNN	Generalized regression neural network
FOA	Fruit fly optimization algorithm
LLRBFNN	Local linear radial basis function neural network
FSA	Fish swarm algorithm
LSTM	Long short-term memory
GA	Genetic algorithm
ML	Machine learning
GD	Gradient descent
MNIST	Mixed National Institute of Standards and Technology
GSA	Gravitational search algorithm
NCL-NN	Negative correlation learning neural network
GOA	Grasshopper optimization algorithm
NFN	Neural fuzzy network
GP	Genetic programming
NN	Neural network
GPU	Graphics processing unit
NNARX	Neural nonlinear auto-regressive exogenous
GSO	Group search optimization
PUNN	Product unit neural network
GWO	Grey wolf optimizer
QRNN	Quantile regression neural network
HS	Harmony search
QNN	Qubit neural network
JA	Jaya algorithm
RaANN	Randomized artificial neural network
MEA	Memetic evolution algorithm
RBFNN	Radial basis function neural network
MH	Meta-heuristic
RBM	Restricted Boltzmann machine
MOO	Multi-objective optimization
RFNN	Recurrent fuzzy neural network
NSGA-II	Non-dominated sorting genetic algorithm
RL	Reinforcement learning
PSO	Particle swarm optimization
RNN	Recurrent neural network
QBA	Quantum-based algorithm
RRNN	Recurrent random neural network
SA	Simulated annealing
SOFNN	Self-organizing fuzzy neural network
SHO	Selfish herd optimization algorithm
SMRN	Single multiplicative recurrent neuron
SI	Swarm intelligence
SAE	Stacked auto encoder
TBO	Trajectory-based optimization
SVM	Support vector machine
TS	Tabu search
WNN	Wavelet neural network
WWO	Water wave optimization

1 Introduction

Artificial Intelligence (AI) was first introduced in the ideas and hypotheses of Gottfried Leibniz [1]. In 1943, McCulloch and Pitts proposed an evolutionary model of the human brain that began research on the artificial neural network (ANN) [2]. ANNs can learn and recognize and solve a wide range of complex problems. Today, ANNs and deep learning (DL) techniques are the most popular and main methods of machine learning (ML) algorithms [3–10]. Figure 1 compares the accuracy of a typical machine learning algorithm and a deep neural network (DNN). As can be seen, if sufficient data and computational power are available, DL techniques perform better (in terms of accuracy) than conventional machine learning approaches [2].

Since 2006, DL has become a popular topic in machine learning. Its position in AI and data science has been shown in Fig. 2 [10]. DL techniques are superior to traditional ML algorithms due to data availability and systems processing power development [10, 11]. In smaller databases and simple applications, traditional ML algorithms perform better because they are easier to implement. This is one of the most important reasons that neural networks and DL techniques had not grown much in the early years [1, 2, 12]. With the advent of the Big Data era, much faster data collection, storage, updating, and management advances have become possible. In addition, the development of GPU has made efficient processing in large data sets. These dramatic advances have led to recent advances in DL techniques [2, 10]. Additionally, reducing the computation time and increasing the convergence process have increased the popularity of these algorithms [3, 4]. Moreover, the position of DL and ANNs in the taxonomy of artificial intelligence approaches has been shown in Fig. 3.

ANNs have been used in various applications, including function approximation [13, 14], classification [15–20], feature selection [21, 22], medical image registration [6], pattern recognition [23–26], data mining [27], signal processing [28], Nonlinear system identification [29, 30], speech processing [31], etc. In addition, different DL methods have been used in various applications, including classification [32–36], prediction [37–39], Phoneme recognition [40], hand-written digit recognition [41–46], etc.

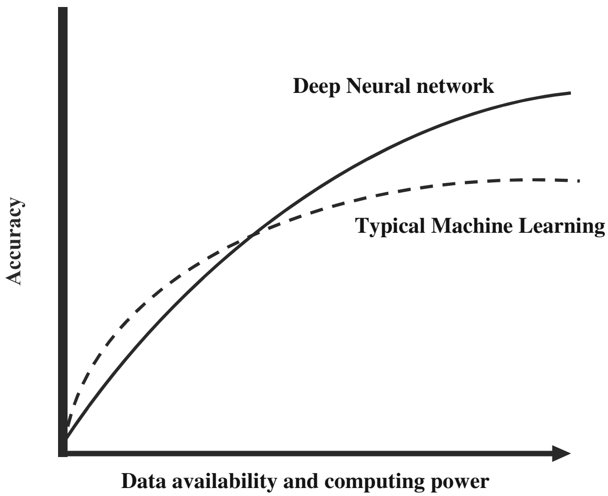

Fig. 1 Comparison of the accuracy of a typical machine learning algorithm and a deep neural network [2]

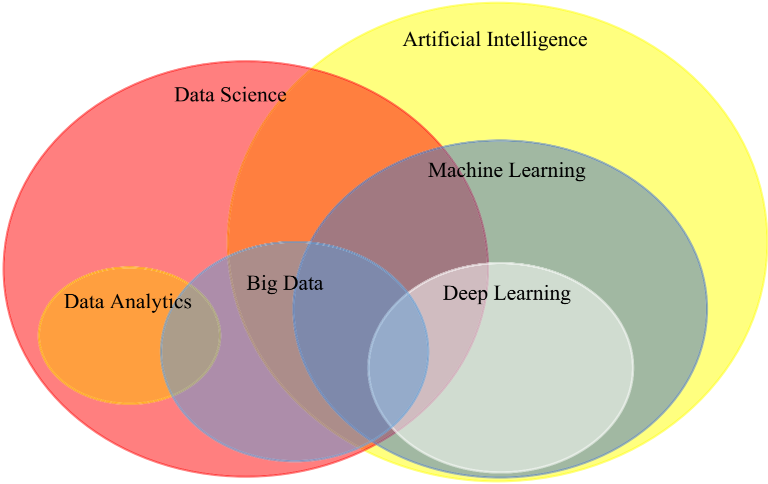

Fig. 2 The position of deep learning in artificial intelligence and data science [10]

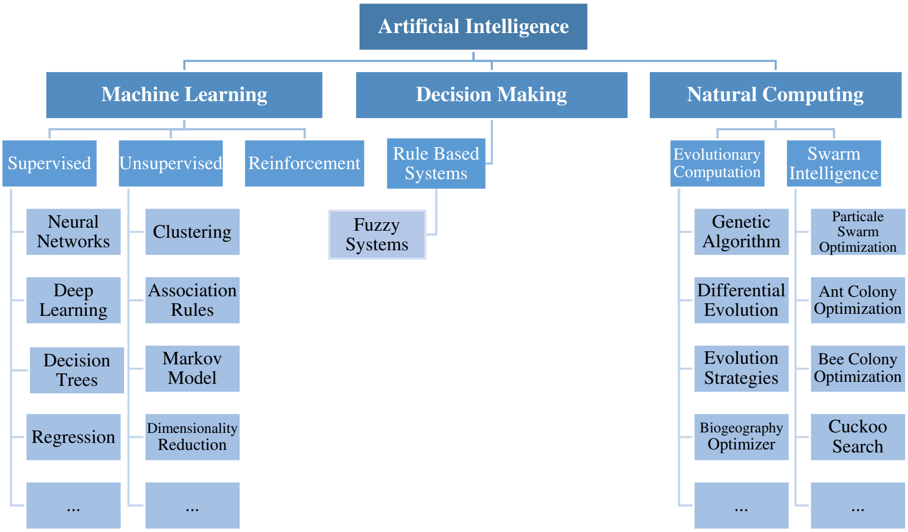

Fig. 3 Taxonomy of artificial intelligence approaches: Machine learning, natural computing, and decision making

Given the importance of using ANNs and DL methods in various applications, identifying weaknesses and improving these algorithms is one of the current issues in machine learning. The learning process of ANNs and DL architectures is considered one of the most difficult machines learning challenges. Over the past two decades, optimizing the structure and parameters of ANNs and DLs has been one of the main interests of researchers [8–10]. Optimization of ANNs and DLs is often considered from several aspects: optimization of weights, hyper-parameters, network structure, activation nodes, learning parameters, learning algorithm, learning environment, etc. [9].

Optimizing weights, biases, and hyper-parameters is one of the most important parts of neural networks and DL architectures. In fact, ANNs and DLs are distinguished by two pillars of structure and learning algorithm. In many past studies, gradient-based methods have been used for architecture training. However, due to the limitations of gradient-based algorithms, the need to use optimization algorithms has been identified [8–10]. For example, in back propagation (BP) learning algorithm, the goal of learning is to optimize the weights and thresholds of the network to minimize the cost function.

In gradient-based learning algorithms, the cost function must be derivative to use BP. This is also one of the weaknesses of gradient-based learning algorithms. Because, in many cases, the activation function (and the cost function) is not derivative. Sigmoid activation functions are commonly used in these algorithms. In the literature, several gradient-based methods, such as Back Propagation (BP) and Levenberg Marquardt (LM) methods, have been developed to teach neural network-based systems [29]. But gradient-based methods have the following major drawbacks.

- For multi-objective cost functions, they may be stuck at local minimums.
- The execution time of these algorithms is very expensive due to the calculation of gradient information with thousands of iterations.
- If there are several local minimums in the problem search space, the learning algorithm reaches error = 0 in the first local minimum. As a result, the learning algorithm converges in the first local minimum and will not achieve the optimal solution. MH algorithms easily escape the local minimum using exploitation and exploration and are a good alternative for gradient-based algorithms.
- In gradient-based learning algorithms, the cost function must be derivative. As a result, the cost function must be continuous. This is also one of the weaknesses of gradient-based learning algorithms. Because, in many cases, the activation function is not derivative. For example, if a step function were used instead of the sigmoid function, all backward calculations in gradient-based learning algorithms would be useless.

At first, Conjugate Gradient Algorithm [47], Newton's Method [48], Stochastic Gradient Descent (SGD) [49], and Adaptive Moment Estimation (Adam) [50] were developed to improve gradient-based learning algorithms, which have better generalizability and convergence than the BP algorithm. However, these methods' neural networks and DL architectures are considered "black boxes" [8]. Because it cannot be interpreted with human intuition. Evolutionary and swarm intelligence algorithms have provided a generalized and optimal network [51–54].

Since training the ANNs and DLs is an NP-hard optimization problem, their structure and parameters optimization using the meta-heuristic (MH) algorithms has been considerably raised. As an optimization problem, MH algorithms formulate the optimal estimation of DL components (such as hyper-parameter, weights, number of layers/neurons, learning rate) [8]. The existence of multiple objectives in optimizing ANNs and DLs, such as error minimization, network generalization, and model simplification, has increased the need for multi-objective MH algorithms. Using MH algorithms to optimize ANNs and DL architectures is still challenging, and more research is needed. Using MH algorithms to train DLs improves the learning process. This increases the accuracy of the algorithm and reduces its execution time.

The rest of the paper is organized as follows: Sect. 2 shows the research methodology. In Sect. 3, first the concept of deep learning models is discussed, then some well-known and state-of-the-art competitive meta-heuristic algorithms are introduced. In Sect. 4, a comprehensive review of the training ANNs and DLs using MH algorithms has been collected. In Sect. 5, the analysis of statistical results from the literature review, challenges and future perspectives are reviewed. Finally, in Sect. 6, the conclusion of this paper is presented.

2 Methodology

This paper has used 440 papers from different journals and publishers in the field of training ANNs and DL architectures (by MH algorithm) for a systematic literature review. First, 627 papers were reviewed, and after reading all the papers, 440 papers entered the next stage. This study systematically searched Google Scholar, Web of Science, and Scopus databases to find related papers. In particular, a thorough search was conducted in Elsevier, IEEE, Springer, Taylor & Francis, John Wiley & Sons, MDPI, Tech Science Press, and other journals. Some conference papers were also selected. In addition, we searched for papers sources to find missing papers. In this paper, only the papers published in English were selected. The following keyword combinations have been used to search for papers:

'Deep learning', 'Artificial neural networks', 'Meta-heuristics', 'Parameters optimization', 'Optimized, 'Training', 'Learning algorithm', 'Deep Autoencoder', 'Adaptive Network Fuzzy Inference System', 'Convolutional Neural Network', 'Deep Boltzmann Machine', 'Deep Belief Network', 'Deep Neural Networks', 'Evolutionary Deep Networks', 'Feed Forward Neural Network', 'Generative Adversarial Network', 'Long Short-Term Memory', 'Machine Learning', 'Radial Basis Function Neural Network', 'Recurrent Neural Network', 'Artificial Bee Colony', 'Ant Colony Optimization', 'Artificial Intelligence', 'Bat Algorithm', 'Biogeography-Based Optimization', 'Chimp Optimization Algorithm', 'Cuckoo Search', 'Differential Evolution', 'Evolutionary Algorithm', 'Evolutionary Computation', 'Evolutionary Deep Learning', 'Evolution Strategy', 'Firefly Algorithm', 'Genetic Algorithm', 'Gravitational Search Algorithm', 'Grasshopper Optimization Algorithm', 'Grey Wolf Optimizer', 'Harmony Search', 'Jaya Algorithm', 'Memetic Evolution Algorithm', 'Multi-objective Optimization', 'Non-dominated Sorting Genetic Algorithm', 'Particle Swarm Optimization', 'Quantum-Based Algorithm', 'Simulated Annealing', 'Swarm Intelligence', 'Trajectory-Based Optimization', 'Tabu Search', and etc.

In this paper, we have tried to collect and discuss all research from the beginning of 1988 to 2022 (September), and therefore 627 articles were selected. The bibliometric tool in this paper was such that first, all papers' titles and the abstract quality of journals based on JCR were reviewed. After this initial review, 187 papers were deleted. Then, the papers that entered the next phase were thoroughly reviewed, and all the discussions and challenges related to this literature review were presented in the next sections.

After analyzing the candidate papers, we found that optimizing the parameters of artificial neural networks and deep learning architectures is a major challenge, and meta-heuristic algorithms are a promising way to solve this challenge. We also noticed that by the mid-2022, there would be a big gap in collecting all papers in this field. Finally, the research questions that need to be answered are as follows:

(1) Why is the optimization of ANNs and DL parameters important?
(2) Which MH algorithms are more used to optimize ANNs and DL architectures?
(3) Which of the ANN and DL parameters are optimized by meta-heuristic algorithms?
(4) Which applications (and dataset) are solved by DLs optimized by meta-heuristic algorithms?

(5) Which ANN and DL architectures are optimized by meta-heuristic algorithms?
(6) What is the effect of using meta-heuristic algorithms to optimize ANNs and DL architectures?
(7) What is the effect of improving meta-heuristic algorithms (and combination of MHs) to optimize ANNs and DL architectures?

3 Background

In the late 1990s, two events created a new challenge in neural networks that marks the beginning of DL today. Long short-term memory (LSTM) was introduced by Hochreiter and Schmidhuber in 1997 and is still one of the most popular DL architectures [55]. In 1998, LeCun et al. developed the first convolutional neural network (CNN), LeNet-5, which yielded significant results in the MNIST dataset [56]. Neither CNN nor LSTM attracted the attention of the large AI community at the time. The last event in the return of deep neural networks (DNNs) was a paper by Hinton et al. in 2006 that introduced deep belief networks (DBN) and produced far better results in the MNIST dataset [57, 58]. After this paper, the renaming of deep neural networks to DL was completed, and a new era in the history of AI began. Figure 4 shows common DL architectures, which are: Long short-term memory (LSTM), Convolutional Neural Networks (CNNs), Deep Belief Networks (DBN), Recurrent Neural Networks (RNN), Deep Boltzmann Machines (DBM), Deep Auto Encoder (DAE), and Deep Neural Networks (DNN).

Much more research is needed to train and optimize the parameters and structure of ANNs and DL architectures. The learning process of ANNs and DLs is one of the most difficult machines learning challenges and has recently attracted the attention of many researchers [8, 10]. Figure 5 shows an example of the evolutionary deep learning architecture (PSO-DCNN) for classification problem.

In recent years, MH algorithms have emerged as a promising method for training ANNs and DLs. The term MH was first introduced in 1986 by Glover [59]. MH methods have become very popular in the last two decades. In designing the MH algorithm, two contradictory criteria are considered: Exploration in the search space and exploitation of the best solutions. In exploration, unsearched areas are visited to ensure that all areas of the search space are searched uniformly. Potential areas are explored more fully in exploitation to find a better solution. Unlike exact methods, MHs solve large-scale problems in a reasonable time. Figure 6 shows the different types of MHs, which include four main categories.

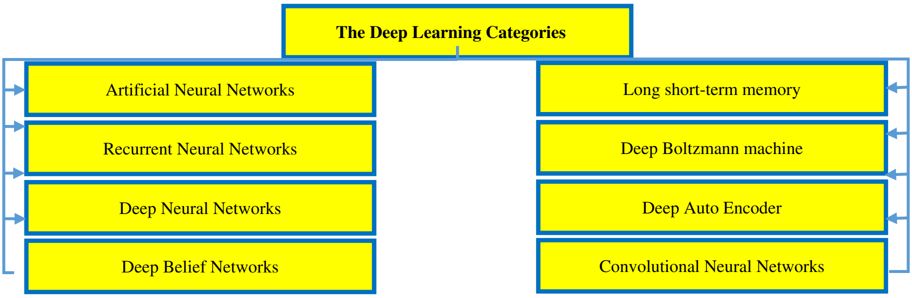

Fig. 4 Common deep learning architectures

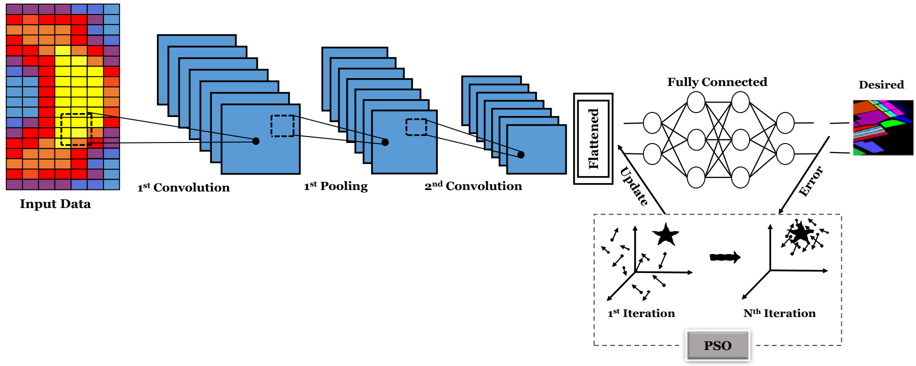

Fig. 5 An example of the evolutionary deep learning architecture (PSO-DCNN) for classification problem

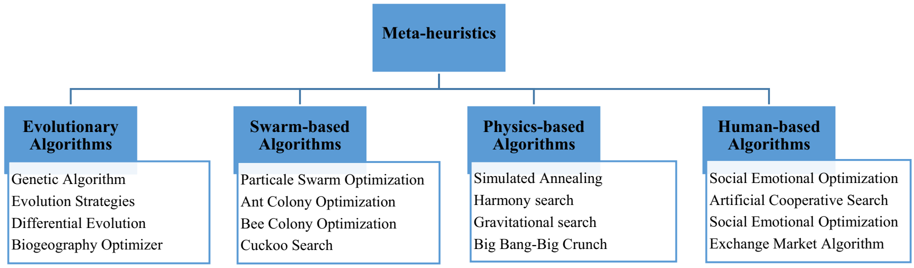

Fig. 6 Different types of meta-heuristic algorithms

Since a few decades ago, a few nature-inspired meta-heuristic algorithms, such as genetic algorithm (GA) [60], ant colony optimization (ACO) [61], particle swarm optimization (PSO) [62], simulated annealing (SA) [63], and differential evolution (DE) [64] have been introduced and used for different optimization problems. Afterward, many studies concentrated on the improvement or adaptation of these MH algorithms for new applications. Other researchers tried to introduce new meta-heuristic algorithms by taking inspiration from nature. Some newer algorithms such as the grey wolf optimization (gwo) [65], black widow optimization (BWO) [66], chimp optimization algorithm (ChOA) [67], red fox optimization (RFO) [68], and gannet optimization algorithm (GOA) [69] are the results of such efforts. Table 1 presents general information about some of the more popular algorithms. In the following, five well-known algorithms called particle swarm optimization (PSO), genetic algorithm (GA), artificial bee colony (ABC), differential evolution (DE), biogeography-based optimization (BBO), and two state-of-the-art competitive algorithms called grey wolf optimization (GWO), and chimp optimization algorithm (ChOA) are introduced.

3.1 Genetic Algorithm (GA)

Genetic algorithm is an exploratory search inspired by Charles Darwin's theory of natural evolution, first introduced by Holland in 1975 [60]. This algorithm reflects the natural selection process in which the best individuals for reproduction are selected to produce offspring.

Table 1 General information of some meta-heuristic algorithms

Authors and references	Algorithm's name and abbreviation	Year
Holland [60]	Genetic algorithm (GA)	1975
Kirkpatrick et al. [63]	Simulated annealing (SA)	1983
Glover [59]	Tabu search (TS)	1986
Srinivas and Deb [70]	NSGA for multi-objective optimization	1994
Eberhart and Kennedy [62]	Particle swarm optimization (PSO)	1995
Dorigo et al. [61]	Ant colony optimization (ACO)	1996
Storn and Price [64]	Differential evolution (DE)	1997
Rubinstein [71]	Cross entropy method (CEM)	1997
Mladenovic and Hansen [72]	Variable neighborhood search (VNS)	1997
Hansen and Ostermeier [73]	CMA-ES	2001
Geem et al. [74]	Harmony search (HS)	2001
Hanseth and Aanestad [75]	Bootstrap algorithm (BA)	2001
Larranaga and Lozano [76]	Estimation of distribution algorithms (EDA)	2001
Pham et al. [77]	Bees algorithms (BA)	2005
Karaboga [78]	Artificial bee colony algorithm (ABC)	2005
Krishnanand and Ghose [79]	Glowworm swarm optimization (GSO)	2006
Haddad et al. [80]	Honey-bee mating optimization (HMO)	2006
Mucherino and Seref [81]	Monkey search (MS)	2007
Atashpaz-Gargari and Lucas [82]	Imperialist competitive algorithm (ICA)	2007
Simon [83]	Biogeography-based optimization (BBO)	2008
Teodorović [84]	Bee colony optimization (BCO)	2009
He et al. [85]	Group search optimizer (GSO)	2009
Yang and Deb [86]	Cuckoo search (CS)	2009
Rashedi et al. [87]	Gravitational search algorithm (GSA)	2009
Kashan [88]	League championship algorithm (LCA)	2009
Kadioglu and Sellmann [89]	Dialectic search	2009
Shah-Hosseini [90]	Intelligent water drops (IWD)	2009
Yang [91]	Firefly algorithm (FA)	2009
Battiti and Brunato [92]	Reactive search optimization (RSO)	2010
Yang [93]	Bat algorithm (BA)	2010
Shah-Hosseini [94]	Galaxy-based search algorithm (GbSA)	2011
Tamura and Yasuda [95]	Spiral optimization (SO)	2011
Alsheddy [96]	Guided local search (GLS)	2011
Rajabioun [97]	Cuckoo optimization algorithm (COA)	2011
Gandomi and Alavi [98]	Krill Herd (KH) algorithm	2012
Civicioglu [99]	Differential search algorithm (DS)	2012
Sadollah et al. [100]	Mine blast algorithm (MBA)	2013
Hatamlou [101]	Black hole (BH)	2013

Table 1 (continued)

Authors and references	Algorithm's name and abbreviation	Year
Gandomi [102]	Interior search algorithm (ISA)	2014
Cheng and Prayogo [103]	Symbiotic organisms search (SOS)	2014
Mirjalili et al. [65]	Grey wolf optimizer (GWO)	2014
Kashan [104]	Optics inspired optimization (OIO)	2015
Kaveh and Mahdavi [105]	Colliding bodies optimization (CBO)	2015
Salimi [106]	Stochastic fractal search (SFS)	2015
Zheng [107]	Water wave optimization (WWO)	2015
Dogan and olmez [108]	Vortex search algorithm (VSA)	2015
Wang et al. [109]	Elephant herding optimization (EHO)	2015
Kashan et al. [110]	Grouping evolution strategies (GES)	2015
Mirjalili [111]	Dragonfly algorithm	2016
Liang et al. [112]	Virus optimization algorithm (VOA)	2016
Mirjalili [113]	Sine cosine algorithm (SCA)	2016
Ebrahimi and Khamehchi [114]	Sperm whale algorithm (SWA)	2016
Mirjalili et al. [115]	Salp swarm algorithm (SSA)	2017
Baykasoğlu and Akpinar [116]	Weighted superposition attraction (WSA)	2017
Mortazavi et al. [117]	Interactive search algorithm (ISA)	2018
Heidari et al. [118]	Harris Hawks optimization (HHO)	2019
Yapici and Cetinkaya [119]	Pathfinder algorithm (PFA)	2019
Kaur et al. [120]	Tunicate swarm algorithm (TSA)	2020
Hayyolalam and Kazem [66]	Black widow optimization (BWO)	2020
Khishe and Mosavi [67]	Chimp optimization algorithm (ChOA)	2020
Braik et al. [121]	Capuchin search algorithm (CapSA)	2021
Talatahari et al. [122]	Crystal structure algorithm (CryStAl)	2021
Połap and Woźniak [68]	Red fox optimization (RFO)	2021
Pan et al. [69]	Gannet optimization algorithm (GOA)	2022
Eslami et al. [123]	Aphid–Ant mutualism (AAM)	2022
Hashim et al. [124]	Honey Badger algorithm (HBA)	2022

This algorithm repeatedly changes the population of individual solutions. In each generation, GA randomly selects individuals from the current population and uses them as parents to produce offspring for the next generation. Over successive generations, the population "evolves" toward an optimal solution. Four phases are considered in a GA.

- *Initial Population* This process begins with a group of chromosomes called a population. Each chromosome is a solution to the problem you want to solve. A chromosome is characterized by a set of variables called genes.
- *Selection* Two pairs of chromosomes (parents) are selected based on their fitness scores. Chromosomes with high fitness have more chance to be selected for reproduction.
- *Crossover* This operator is the most significant step in a GA algorithm. For each pair of parents to be mated, a crossover point is randomly selected from within the genes. Offspring are created by exchanging the genes of parents. The crossover operator is applied

Chromosome	W_1	W_2	…	W_n	B_1	B_2	…	B_m

Fig. 7 Chromosome definition in GA

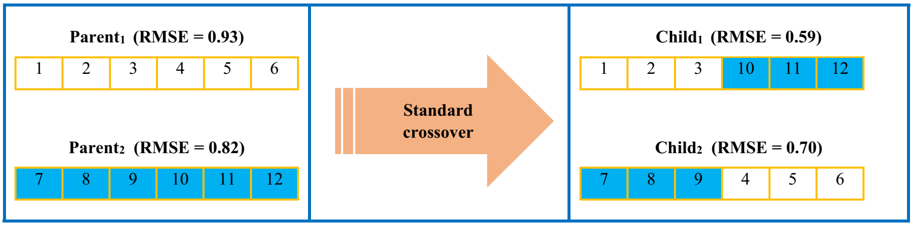

Fig. 8 An example of single point crossover

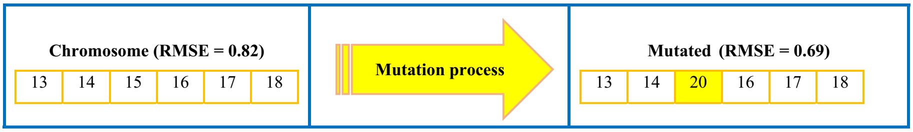

Fig. 9 Example of the mutation operator in GA

to improve the exploitation of algorithm. This operator actually searches the space around a chromosome.

- *Mutation* In some newly formed offspring, some of their genes can be subjected to a mutation. The mutation operator is applied to enhance exploration.

Today in many applications, GA is used to train the deep learning architectures such as convolutional neural network (GA-CNN). In this proposed architectures, GA optimizes the weights and biases of the CNN. In the following, GA modeling for this problem is presented. For GA modeling, one of the main tasks is to define a solution in the form of a chromosome. Figure 7 shows the definition of a chromosome in GA.

Figure 8 shows the single point crossover operator of standard GA. As can be seen, in a single-point crossover, only two chromosomes are combined. Figure 9 illustrates the mutation process of GA.

3.2 Differential Evolution (DE)

Differential evolution (DE) is a global optimization algorithm developed by Storn and Price in the year 1997 [64]. Similar to other popular approaches, such as genetic algorithm and evolutionary algorithm, the differential evolution starts with an initial population of candidate solutions. These candidate solutions are iteratively improved by introducing crossover, mutation, and selection into the population, and retaining the fittest candidate solutions. Due to its several competitive advantages, DE is one of the most popular MH algorithm used by researchers and practitioners to tackle a diverse set of real-world applications. First, the implementation of DE is simpler than most other MHs. This feature enables those practitioners who may not have strong coding skills to make simple adjustments to the DE coding to solve

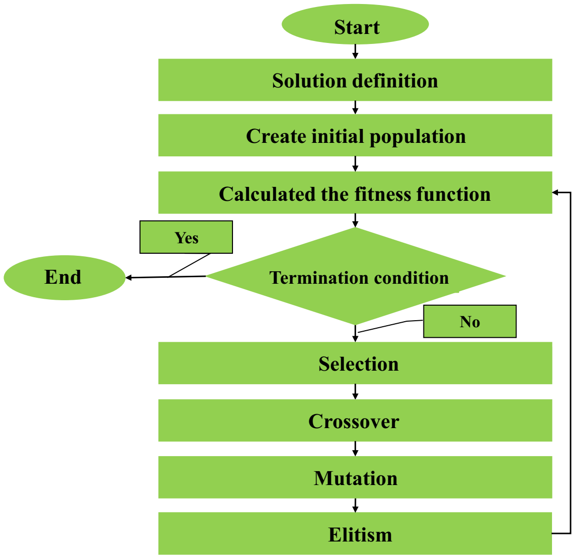

Fig. 10 The flowchart of DE algorithm

problems. Second, despite its simplicity, DE can show a more promising optimization ability than other MHs in solving different types of optimization problems such as nonlinearity and multimodality. Third, various DE algorithms have appeared as the top three best-performing optimizers in most CEC competitions since 2005. Figure 10 shows the flowchart of the DE algorithm.

3.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) algorithm is one of the most important intelligent optimization algorithms in the field of Swarm Intelligence. This algorithm was introduced by Kennedy and Eberhart in 1995, inspired by the social behavior of animals such as fish and birds that live together in small and large groups. PSO is suitable for a wide range of continuous and discrete problems and has performed very well in different optimization problems [62].

In PSO, all possible solutions are mapped to corresponded particles, and every particle is assigned an initial velocity that deputes a position change. For calculating the next velocity of the particles in the solution space, an optimization function is utilized. Particle velocity is made of three main movements: a) the percentage of the previous movement's continuation, b) the movement toward the best personal experience, and c) the movement toward the best global experience. Equations (1) and (2) are respectively expressing the update of velocity and position of the particles.

$$V_{id}(\mathrm{t}+1) = V_{id}(\mathrm{t}) + \text{firand}(0, '_1)(P_{id}(t) - X_{id}(t)) + \text{firand}(0, '_2)(P_{gd}(t) - X_{id}(t)) \quad (1)$$

$$X_{id}(\mathrm{t}+1) = X_{id}(\mathrm{t}) + V_{id}(\mathrm{t}+1) \quad (2)$$

3.4 Artificial Bee Colony (ABC)

Artificial bee colony (ABC) is a swarm based meta-heuristic algorithm that was introduced by Karaboga in 2005. ABC was inspired by the intelligent search behavior of honey bees [78]. In ABC algorithm, the colony contains three types of artificial bees (Fig. 11):

- *Scout bees* Solutions that are randomly generated to discover new spaces are called scout bees. Scout bees are responsible for exploring the search space.
- *Employed bees* A number of scout bees with good fitness function become employed bees. Employed bees are responsible for advertising quality food sources.
- *Onlooker bees* The onlooker bees are responsible for searching the neighborhood for employed bees. Onlooker bees receive information about food sources and search around these sources. The role of these bees is both exploitation and exploration of algorithm.

In ABC, scout bees randomly discover a population of initial solution vectors and then repeatedly improve them by onlooker and employed bees (using neighbor search method to move towards better solutions while eliminating poor solutions). In general, ABC uses two main methods (neighbor search and random search) to get the optimal answer: Random search by scout and onlooker bees and neighbor search by employed and onlooker bees. In ABC, each candidate answer indicates the position of food source, and the quality of the nectar is used as a fitness function. In this algorithm, first, all initial populations are explored by scout bees. Scout bees with best fitness functions are selected as the employed bees. Employed bees exploit the solution positions and then onlooker bees are created. The higher the quality of the employed bee, the more onlooker bees will be created around it. The onlooker bee also select new food positions (using the employed bee information) and exploit around these positions. In the next step, random scout bees are created to find new

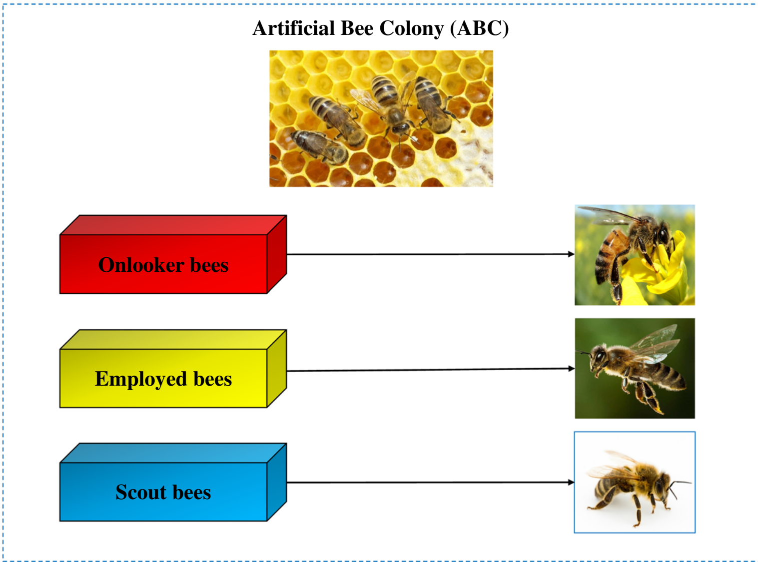

Fig. 11 Three types of artificial bees in ABC

random food positions. ABC algorithm can be formulated as Eq. (3)-(5).

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{3}$$

$$V_{ij} = X_{ij} + \varphi_{ij}(X_{ij} - X_{kj}) \tag{4}$$

$$X_L^j = X_{min}^j + rand(0, 1)(X_{max}^j - X_{min}^j) \tag{5}$$

where.

P_i = Probability of selecting employed bees by onlooker bees.
fit_i = Fitness function of the i^{th} solution.
V_{ij} = Onlooker bee.
X_L^j = Scout bees.
X_{min}^j = Low limit of search space.
X_{max}^j = High limit of search space, SN = Number of employed bees.
$i \in$ {1, 2, …, SN}.
j = Dimension ∈ {1, 2, …, D}.
k = Onlooker bee number.
φ_{ij} is the random number ∈ [0, 1]
L = Scout bee number.

3.5 Biogeography-Based Optimization (BBO)

Biographical-based optimization is a population-based evolutionary algorithm first proposed by Dan Simon in 2008 [83]. The answer in BBO is called habitat and habitat is considered as a vector of its habitant. In addition, the value of each habitat is defined by the habitat suitability index (HSI). The high value of HSI shows high fitness function of habitat. Three main operators of BBO include migration, mutation and elitism. In BBO, each habitat has its own emigration rate, immigration rate, and mutation rate. The emigration ($\mu_j(k)$) rate and immigration rate ($\lambda_j(k)$) are defined as Eq. (6) and Eq. (7).

$$\mu_j(k) = E \times (\frac{k(j)}{N}) \tag{6}$$

$$\lambda_j(k) = I \times (1 - \frac{k(j)}{N}) \tag{7}$$

In which, $k(j)$ represents the rank of the jth habitat after sorting accordance to their HSI and N is the highest rank in the total habitat (population size). The rank $k(j)$ is related to the habitat suitability index (fitness function). In addition, E represents the highest emigration rate and I represents the highest immigration rate. Migration, mutation and elitism are the main operators of this algorithm. By assuming H_i as the host habitat and H_j as the guest habitat, the migration process for the standard BBO will be as the Eq. (8):

$$H_i(SIVs) \leftarrow H_i(SIVs) + H_j(SIVs) \tag{8}$$

According to the Eq. (8), the host habitat (selected based on the immigration rate and roulette wheel method) receives information only from the guest habitat (selected based on the emigration rate and roulette wheel method) and itself.

3.6 Grey Wolf Optimization (GWO)

GWO is a swarm-based MH algorithm inspired by the the gray wolf's hunting policies [65]. GWO divide the population into four levels: alpha, beta, delta, and omega. Alphas are the leaders that make decisions about living, hunting, and moving wolfs, while the beta act as an advisor to the alpha. The delta is responsible for warning when there is danger and protecting the pack, providing food and caring for sick or injured wolves. In the end, Omega is the last wolve that has to obey leaders. They follow four phases: hunting, searching, encircling, and then attacking the prey. GWO is one of the state-of-the-art competitive MH algorithm, which has attracted great attention of researchers. GWO is simple to set parameters, flexible and has a good trade-off between exploration and exploitation.

3.7 Chimp optimization Algorithm (ChOA)

ChOA algorithms is one of the new MH algorithm introduced by Khishe and Mosavi in 2020. ChOA is inspired by the chimps' movement in group hunting and their sexual motivations [67]. In the ChOA, prey hunting is utilized to reach the optimal solution in the optimization problem. ChOA divides hunting into four main phases: driving, blocking, chasing, and attacking. In the first, ChOA is initialized by the generating a random chimps' population. Chimps are then randomly classified into four groups: attacker, chaser, barrier, and driver. In order to model driving and chasing the prey, Eqs. (9)–(13) have been proposed.

$$\boldsymbol{d} = \left|\boldsymbol{c}.\boldsymbol{X}_{prey}(t) - \boldsymbol{m}.\boldsymbol{X}_{chimp}(t)\right| \tag{9}$$

$$\boldsymbol{X}_{chimp}(t+1) = \boldsymbol{X}_{prey}(t) - \boldsymbol{a}.\boldsymbol{d} \tag{10}$$

$$\boldsymbol{a} = 2.\boldsymbol{f}.\boldsymbol{r}_1 - \boldsymbol{f} \tag{11}$$

$$\boldsymbol{c} = 2.\boldsymbol{r}_2 \tag{12}$$

$$\boldsymbol{m} = \boldsymbol{Chaotic\_value} \tag{13}$$

where, $\boldsymbol{X}_{prey}$ is the prey position vector, $\boldsymbol{X}_{chimp}$ denote the chimp position vector, t present the current iteration, $\boldsymbol{a}$, $\boldsymbol{c} and \boldsymbol{m}$ are the coefficient vectors, $\boldsymbol{f}$ is the dynamic vector $\in [0, 2.5]$, $\boldsymbol{r}_1 and \boldsymbol{r}_2$ are the random vectors $\in [0, 1]$, and $\boldsymbol{m}$ denote a chaotic vector.

The chimps first detect the prey's position in the hunting step using driver, blocker, and chaser chimps. In the exploitation process, the hunting process is done by attackers. For this purpose, the prey's position is estimated by the attacker, barrier, chaser, and driver chimps, and other chimps update their position through the prey. This process is formulated as Eqs. (14)–(16).

$$\begin{array}{ll} \boldsymbol{d}_{Attacher} = |\boldsymbol{c}_1.\boldsymbol{X}_{Attacher} - \boldsymbol{m}_1.\boldsymbol{X}|, & \boldsymbol{d}_{Barrier} = |\boldsymbol{c}_2.\boldsymbol{X}_{Barrier} - \boldsymbol{m}_2.\boldsymbol{X}| \\ \boldsymbol{d}_{Chaser} = |\boldsymbol{c}_3.\boldsymbol{X}_{Chaser} - \boldsymbol{m}_3.\boldsymbol{X}|, & \boldsymbol{d}_{Driver} = |\boldsymbol{c}_4.\boldsymbol{X}_{Driver} - \boldsymbol{m}_4.\boldsymbol{X}| \end{array} \tag{14}$$

$$\begin{array}{ll} \boldsymbol{X}_1 = \boldsymbol{X}_{Attacher} - \boldsymbol{a}_1(\boldsymbol{d}_{Attacher}), & \boldsymbol{X}_2 = \boldsymbol{X}_{Barrier} - \boldsymbol{a}_2(\boldsymbol{d}_{Barrier}) \\ \boldsymbol{X}_3 = \boldsymbol{X}_{Chaser} - \boldsymbol{a}_3(\boldsymbol{d}_{Chaser}), & \boldsymbol{X}_4 = \boldsymbol{X}_{Driver} - \boldsymbol{a}_4(\boldsymbol{d}_{Driver}) \end{array} \tag{15}$$

$$\boldsymbol{X}(t+1) = \frac{\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3 + \boldsymbol{X}_4}{4} \tag{16}$$

where, $\boldsymbol{X}_{Attacher}$ denotes the best search agent, $\boldsymbol{X}_{Barrier}$ is the second-best search agent, $\boldsymbol{X}_{Chaser}$ presents the third-best search agent, $\boldsymbol{X}_{Driver}$ is the fourth-best search agent, and $\boldsymbol{X}(t+1)$ is the updated position of each chimp.

Also, to set up the exploration process, $\boldsymbol{a}$ parameter is applied such that $\boldsymbol{a} > 1$ and $\boldsymbol{a} < -1$ is the cause of diverging chimps and preys. As well, $\boldsymbol{a}$ parameter with the values between + 1 and − 1, help the chimps and preys to be converged and will lead to improved exploitation. In addition, $\boldsymbol{c}$ parameter helps the algorithm to have the exploration process. Finally, all chimps attack their prey to achieve social rights (sexual incentive) after prey hunting regardless of their duties. In order to formulate social behavior, chaotic maps are used as Eq. (17).

$$\boldsymbol{X}_{chimp}(t+1) = \begin{cases} \boldsymbol{X}_{prey}(t) - a.dif \mu < 0.5 \\ Chaotic\_value if \mu \geq 0.5 \end{cases} \tag{17}$$

Where, μ is the random number $\in [0, 1]$

3.8 Memetic Algorithms (Hybridization)

It is complicated to find the best possible solution in the search space in large-scale optimization problems. Moreover, changing algorithm variables does not have much influence on the algorithm convergence. Therefore, for massive dataset with high complexity, even if the researchers have determined accurate initial parameters, the algorithm will not be able to perform adequate exploration and exploitation. Consequently, to achieve comprehensive global and local searches, we need to apply powerful operators to make better exploration and exploitation. MH algorithms can be combined with others and overcome this problem by using the advantages and operators of other algorithms [125]. Despite promising results achieved by MHs over the past years, many successful attempts have been made that do not pursue a single inspiration from nature but compound various MHs exploiting their complementarity. This is particularly important for challenging optimization applications where combination methods show promising performance, leading to further intensification of the research. Generally, High-level hybridization of MHs is achieved by running algorithms in a sequence where all factors changed by one MH are transferred to the other algorithm [125]. According to the literature review, most hybridization models are designed for specific optimization problem, including clustering, feature selection, and image segmentation. Since modelling a hybrid model that would be able to improve more than one MH is challenging, available solutions mostly use two competitive algorithms to an optimization problem. In recent decades, researchers have utilized a combination of algorithms to improve the performance of the optimization process.

3.9 Modification of MH (Devoted Local Search and Manipulating the Solutions Space)

The increasing discovery of alternative methods to solve optimization problems makes it necessary to parallelize and modify available algorithms. Achieving a suitable solution using a MH algorithms may need a long runtime, iterations, or population. The first one is to use the neighborhood search method in order to minimize the exploration of the solution space. In addition, powerful CPU can affect the convergence speed of the MH algorithm and therefore work more efficiently. In the proposed neighborhood search approach, smaller populations called groups may formed. Suppose the number of computer cores is specified at the beginning of the algorithm. In comparison with the standard version of MH algorithms, an initial population consisting of N individuals is generated randomly. From this population, suitable individuals are selected. Each individual in population will be the best adapted solution in the smaller group that will be created under his leadership. The second proposed approach involves manipulating the solutions space to minimize the number of calculations. In this proposition, the multi-threading approach plays a big role because dividing the space and selecting the best areas does not cost extra. In addition, the third proposed approach is the combination of the previous two methods. While the proposed approach of parallelization and manipulation of solution space improves the performance of classical algorithms, they are so flexible that can be improved with different ideas. In addition, it achieves better results in different applications [126].

4 Review of the Training DL and AANs by MH Algorithms

This section provides an overview of the optimization of neural networks and DL architectures using MH algorithms. The review of papers is divided into two parts: ANN optimization and DL optimization.

4.1 Review1: Training the AANs by MH Algorithms

This section provides a comprehensive overview of the optimization of different types of ANNs using MH algorithms. Optimization of ANNs is often considered from several aspects: optimization of weights, hyper-parameters, network structure, activation nodes, learning parameters, learning algorithm, learning environment, etc.

Eberhart and Kennedy [62] used the PSO algorithm to optimize the weights of an MLPNN. The proposed architecture performed very well on a benchmark data set. Storn and Price [64] used a differential evolution algorithm to optimize the weights of an FFNN. Experiments on the nonlinear optimization problem indicated the superiority of the proposed DE-FFNN algorithm. PSO algorithm was used by Chunkai et al. [127] to optimize the weights and architecture of MLPNN. This hybrid approach was introduced to model the quality estimation of a product. The results showed that the performance of PSO-MLPNN is better than other algorithms. Li et al. [128] used the genetic algorithm to train the parameters and weights of an ANN. The proposed architecture (GA-ANN) showed good performance for the pollutant emissions problem.

Leung et al. [129] used the improved genetic algorithm (IGA) to optimize the architecture and weights of an ANN. This study compared the proposed architecture (IGA-ANN) with other architectures and presented better results. Meissner et al. [130] used an improved PSO algorithm to optimize the number of neurons, parameters, and weights of an ANN. The developed architecture showed good results in benchmark datasets. Geethanjali et al. [131] used the PSO algorithm to train the ANN (MLFFNN). The results showed that the PSO-MLFFNN architecture was more accurate and faster than the BP- MLFFNN architecture. Yu et al. [132] used PSO and DPSO algorithms to optimize the architecture and parameters (weight and bias) of a three-layer FFANN network. The proposed algorithm was named ESPNet. A self-adaptive evolutionary strategy was used to improve PSO and DPSO. Experimental results from two real-world problems show that ESPNet can generate compact neural networks with good generalizability.

Khayat et al. [133] used GA and PSO algorithms to optimize the weights of a SOFNN. The results showed that the optimized SOFNN architecture based on GA and PSO performs well. Lin and Hsieh [134] used the improved PSO algorithm to optimize the weights of a three-layer neural network. The proposed approach provided good performance for the classification data. Cruz-Ramírez et al. [135] used the Pareto Memetic Differential Evolution Algorithm (MPDA) to optimize the structure and weights of a neural network. The proposed approach performed well in benchmark problems. Subudhi and Jena [29] used the combination of the memetic differential evolution (MDE) algorithm and BP algorithm (DEBP) to train a multilayer neural network to identify a nonlinear system. DEBP performance was compared with six other algorithms such as Back Propagation (BP), Genetic Algorithm (GA), PSO, DE, Back Propagation genetic algorithm (GABP), and Back Propagation Particle Swarm Optimization (PSOBP). The results of different algorithms showed that the proposed DEBP has better identification compared to other cases.

Malviya and Pratihar [136] used PSO, BP, and two clustering algorithms (including Fuzzy C-means) to train the RBFNN and MLFFNN networks for the MIG welding process problem. In this research, connection weights and learning parameters are optimized. Zhao and Qian [137] used the CPSO algorithm to optimize the weights and architecture of a three-layer FFNN. The performance of CPSO-FFNN was compared with the existing architectures in the research literature, and the results showed the superiority of the proposed architecture. Green II et al. [138] used the CFO algorithm to optimize the weights of an ANN. The performance of the CFO was compared with the PSO algorithm, which shows the superiority of CFO-NN.

Vasumathi and Moorthi [139] used the PSO algorithm to optimize the weights of an ANN. The results showed that the proposed PSO-ANN architecture performs well in the harmonic estimation problem. Yaghini et al. [140] used a combination of the improved particle swarm optimization (IOPSO) and the BP algorithm to train an ANN. The developed architecture was implemented on eight benchmark datasets. IOPSO-BPA-ANN also performed better than the other 10 algorithms. Dragoi et al. [141] used the differential evolutionary self-adaptation algorithm (SADE) to optimize the weights, architecture, and learning parameters of an ANN. The developed approach for the aerobic fermentation process was proposed and presented good results. Ismail et al. [142] used a combination of PSO and BP algorithms to train the product unit neural network (PUNN). The PSO-BP-PUNN architecture performed better than the PSO-PUNN and BP-PUNN architectures.

Das et al. [143] used the PSO algorithm to train ANN. In this study, all four parameters of weight, number of layers, number of neurons and learning parameters were optimized simultaneously. According to the results, the PSO-ANN architecture performed better than other architectures in the literature. Mirjalili et al. [144] used the BBO algorithm to optimize the weights of an MLPNN for classification and function approximation problems. They compared the BBO algorithm with five other metaheuristic algorithms and the BP and ELM algorithms. BBO results were better than other algorithms in terms of accuracy and convergence speed. Jaddi et al. [145] used the improvement of the bat algorithm to optimize an ANN. Where both the ANN structure and the network weights are optimized. Statistical analysis showed that the bat algorithm with Ring and Master-Slave strategies for the classification problem performed better than other methods in the literature.

Jaddi et al. [146] used the improved bat algorithm (MBA) to optimize the weights, architecture, and active neurons of an ANN. The hybrid algorithm showed high performance in six classification problems, two-time series problems and one real-world problem. González et al. [147] used the fuzzy gravitational search algorithm (FGSA) to train a neural network's modules, layers and nodes. The proposed FGSA-NN architecture was implemented for the pattern recognition problem and provided acceptable results. Gaxiola et al. [148] used particle swarm optimization and a genetic algorithm to optimize the weights of type-2 fuzzy inference systems. The developed architectures were implemented on time series benchmark datasets. According to the results, NNT2FWGA and NNT2FWPSO algorithms performed better than

NNT2FW. Karaboga and Kaya [149] used the hybrid artificial bee colony algorithm (aABC) to train ANFIS. The performance of aABC-ANFIS was compared with 14 other architectures on four nonlinear dynamic systems, which showed its superiority in accuracy.

Jafrasteh and Fathianpour [150] used an improved artificial bee colony algorithm (SPABC) to train the LLRBF neural network. The results of the proposed algorithm were compared with six other MH algorithms that show the superiority of SPABC-LLRBFNN. Khishe et al. [19] used the improved migration model of the biogeography-based optimization to optimize the weights and biases of an MLPNN. They developed the exponential-logarithmic migration model to improve BBO performance. Additionally, the performance of the proposed algorithm was compared with six other MH algorithms for sonar data classification, which showed the superiority of IBBO-MLPNN. Ganjefar and Tofighi [151] used a combination of GA and GD algorithms to train an ANN. The proposed HGAGD-NN approach has yielded good results for several benchmark problems.

Aljarah et al. [152] used the whale optimization algorithm (WOA) to train the weights of an MLPNN. They implemented the proposed WOA-MLP algorithm on 20 benchmark problems, which produced better accuracy and speed than the BP, GA, PSO, ACO, DE, ES, and PBIL algorithms. Heidari et al. [153] used the grasshopper optimization algorithm (GOA) to train an MLPNN. The performance of GOA-MLPNN was evaluated with eight other algorithms on five medical identification classification datasets. Finally, the proposed GOA-MLPNN algorithm gave better results in different criteria. Hadavandi et al. [154] proposed an MLPNN simulator based on the gray wolf optimizer (GWO) to predict the tensile strength of Siro-Spun yarn. The gray wolf optimizer algorithm was applied to train the neural network weights. Finally, proposed hybrid architecture GWO-MLPNN performed better than a traditional learning-based neural network (BP-MLPNN).

Haznedar and Kalinli [155] used the SA algorithm to train an ANFIS. The SA-ANFIS architecture was compared with GA, BP algorithms and various architectures from the research literature, which showed the superiority of SA-ANFIS. Pham et al. [156] used biogeography-based optimization to optimize the weights and parameters of an MLPNN to predict the soil composition coefficient. This study used BP-MLPNN, RBFNN, Gaussian Process (GP) and SVR algorithms to compare with BBO-MLPNN. According to the results, the BBO-MLPNN algorithm excelled in three criteria: RMSE, MAE and correlation coefficient. Han et al. [157] used the improved mutation model of the DE algorithm to optimize the neural network. The DE-BPNN model has been implemented to predict the performance of pre-cooling systems, which has yielded far better results than other networks.

Rojas-Delgado et al. [158] used particle swarm optimization (PSO), firefly algorithm (FA), and cuckoo search (CS) to train the ANN. The various neural network architectures trained by meta-heuristic algorithms were implemented on six benchmark problems that performed very well compared to traditional methods. Khishe and Mosavi [159] used the chimp optimization algorithm to optimize the weights and biases of an MLPNN. In that study, the performance of the MLPNN-ChOA algorithm was compared with the performance of IMA, GWO and a hybrid algorithm on the underwater acoustic dataset classification problem, which showed the superiority of the MLPNN-ChOA. Wang et al. [160] used the PSO and CA algorithms to optimize the neural network weights. The combined particle swarm optimization (HPSO) algorithm was first developed in that research. The HPSO algorithm was combined with CA, and finally, the HPSO-CA algorithm was implemented for network training (HPSO-CA-ANN). The developed algorithm and five other MH algorithms were implemented on 15 benchmark datasets that performed better than the others.

Al-Majidi et al. [161] used the PSO algorithm to optimize the weights and architecture of FFNN. The results showed that the optimized FFNN architecture based on the PSO accurately predicts the maximum power point. Ertuğrul [54] used the differential evolution algorithm (DE) to optimize the nodes and learning parameters of RaANN. The results showed that the differential evolution algorithm for 48 synthetic datasets performed better than other methods. Ansari et al. [162] used the magnetic optimization algorithm (MOA) & PSO to optimize the weights of the back-propagation neural network. According to the results, the proposed approach (MOA-BBNN) performed well in the bankruptcy prediction problem.

Zhang et al., [163] used the chicken swarm optimization (CSO) algorithm to optimize the weights, biases, and number of layers of the Elman neural network (ENN). According to the results, the proposed hybrid approach (CSO-ENN) performed well in the Air pollution forecasting. Also, the performance of the proposed hybrid architecture has been better than other algorithms. Li et al., [164] used the biogeography-based optimization (BBO) algorithm to optimize the weights of MLPNN for medical image classification. The results showed that the proposed hybrid architecture (BBO-MLPNN) performs better than the other original architectures.

Table 2 summarizes the above research as well as many other studies. As can be seen, for each research, the author's name, year of publication, type of neural network, optimized components in the network, type of MH algorithm used, application and data set used are listed. In the following, for a more comprehensive review, some statistical analysis of the research collected in Table 2 is presented.

4.1.1 Investigation of Optimized Components in ANNs

As an optimization problem, MH algorithms formulate the optimal estimation of ANN components (such as weights, number of layers, number of neurons, learning rate, etc.). This section examines the abundance of MH use for optimized components in neural networks (according to the papers in Table 2). Figure 12 shows the relative abundance of research on optimized components in ANNs using MH algorithms.

As shown in Fig. 12, in 221 studies (69%), weights and biases have been adjusted using MH algorithms, which shows a high percentage. In 47 studies (14%), the number of neurons in the layers has been adjusted using MH algorithms. Moreover, in 22 studies (7%), the number of layers in the neural network has been adjusted. Finally, in 31 studies (10%), learning parameters, learning algorithms or activation functions have been adjusted. Figure 13 also shows the relative abundance of research in the simultaneous optimization of two components of ANNs.

As can be seen in Fig. 13, in 15 studies, weights and layers have been adjusted simultaneously. In 28 studies, weights and neurons; in 15 studies, weights and learning parameters; in 14 studies, the number of layers and neurons; in 6 studies, the number of layers and learning parameters; and in 14 studies, the number of neurons and learning parameters have been adjusted simultaneously. Figure 14 shows the relative abundance of research in the simultaneous optimization of three components of ANNs. As can be seen, in 6 studies, weights, the number of neurons and learning parameters have been adjusted simultaneously. In 7 studies, weights, number of layers and number of neurons; in 2 studies, weights, number of layers and learning parameters; in 5 studies, number of layers, number of neurons and learning parameters were adjusted simultaneously. According to Table 2, in only one study [143], all four neural network components were adjusted simultaneously. Therefore, little research has been done in this area.

Table 2 A summary of meta-heuristic algorithms developments for training/optimization of ANNs

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Engel [165]	FFNN	✓				Simulated annealing (SA)	The parity and "clump-recognition" problems
Montana and Davis [166]	FFNN					Genetic algorithm (GA)	Sonar data from arrays of underwater acoustic receivers
Whitley et al. [167]	FFNN	✓				Genetic algorithm (GA)	Benchmark problems for Training NN
Belew et al. [168]	FFNN				✓	Genetic algorithm (GA)	Benchmark optimization problems and classification
Kitano [169]	ANN	✓		✓		Genetic algorithm (GA)	Benchmark optimization problems and classification
Eberhart and Kennedy [62]	MLPNN	✓				Particle swarm optimization (PSO)	Systematic benchmark optimization problems
Battiti and Tecchiolli [170]	FFNN	✓				Reactive tabu search (RTS) algorithm	Training sub-symbolic systems
Storn and Price [64]	FFNN	✓				Differential evolution (DE) algorithm	Non-linear optimization problems
Yao and Liu [171]	FFNN	✓		✓		Evolutionary programming (EP)	The Parity and Medical Diagnosis Problems
Sexton et al. [172]	FFNN	✓				Tabu search (TS)	Mackey–Glass chaotic time series & Benchmark problems
Sexton et al. [173]	FFNN	✓				Simulated annealing (SA)	Monte Carlo study on seven test functions

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Chunkai et al. [127]	MLPNN	✓	✓			Particle swarm optimization (PSO)	Modelling product quality estimator problem
Arifovic and Gencay [174]	FFNN	✓		✓		Genetic algorithm (GA)	The long-term behavior of dissipative systems
Alvarez [175]	FFNN				✓	Genetic programming (GP)	The problem domain of time series prediction
Li et al. [128]	ANN	✓			✓	Genetic algorithm (GA)	Human supervisory control, pollutant emission
Sarkar and Modak [176]	FFNN	✓				Simulated annealing (SA) algorithm	Nonlinear optimal control problems
García-Pedrajas et al. [177]	ANN			✓		Cooperative coevolution	Three real problems of classification
Ilonen et al. [178]	FFNN	✓				Differential evolution (DE) algorithm	Continuous optimization problems
Leung et al. [129]	FFNN	✓			✓	Improved genetic algorithm (IGA)	Some benchmark optimization functions
Augusteijn and Harrington [179]	FFNN				✓	Evolutionary programming (EP)	Four benchmark classification problems
Abraham [180]	ANN	✓		✓	✓	Evolutionary algorithm & meta-learning evolutionary	Three different well-known chaotic time series

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Lahiri and Chakravorti [181]	ANN		✓	✓	✓	Genetic algorithm (GA)	Electrode-spacer contour optimization
Shen et al. [182]	MLFFNN	✓		✓		Particle swarm optimization (PSO)	QSAR studies of bioactivity of organic compounds
Kim et al. [183]	FFNN	✓				Genetic algorithm (GA)	Mathematical optimization and set covering problem
Chatterjee et al. [184]	FNN	✓				Particle swarm optimization (PSO)	Optimization voice-controlled robot systems
Feng et al. [185]	FFNN	✓				Guaranteed convergence PSO (GCPSO)	Noise Identification and Classification Problem
Da and Xiurun [186]	FFNN	✓				Modified PSO with simulated annealing (PSOSA)	Triaxial compression tests (rock engineering)
Salajegheh and Gholizadeh [187]	RBF			✓		Improved genetic algorithm (IGA)	25-bar space tower,-bar grid space dome,
Tsai et al. [188]	FFNN	✓		✓		Hybrid Taguchi-genetic algorithm (HTGA)	Forecasting the sunspot numbers
García-Pedrajas et al. [189]	ANN			✓		Genetic algorithm (GA)	25 real-world optimization problems
Meissner et al. [130]	ANN	✓		✓	✓	Optimized particle swarm optimization (OPSO)	Benchmark datasets
Ye et al. [190]	FFNN	✓				Tabu search (TS)	Several typical non-linear optimization functions

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Socha and Blum [191]	FFNN	✓				Ant colony optimization (ACO) algorithm	Discrete optimization problems
Lin et al. [192]	MLFFNN	✓				Particle swarm optimization (PSO)	Application in QSAR studies of bioactivity
Ulagammai et al. [193]	WNN	✓				Bacterial foraging technique (BFT)	Identification of the non-linear characteristics of power system
Zhang et al. [194]	FFNN	✓				Hybrid particle swarm optimization (HPSO)	Three bits parity problem
Yu et al. [132]	3LFFANN	✓	✓			Discrete particle swarm optimization (DPSO) & PSO	Two real-world problems
Geethanjali et al. [131]	MLFFNN	✓				Particle swarm optimization (PSO)	Modeling power transformers problems
Lin et al. [195]	FLNFN	✓				Cooperative particle swarm optimization (CPSO)	Prediction Applications
Tsoulos et al. [196]	FFNN	✓		✓		Grammatical evolution (GE)	9 known classification and 9 known regression problems
Goh et al. [197]	FFNN			✓		Microhybrid genetic algorithm (μHGA)	Real-world medical data sets
Lin and Hsieh [134]	3LNN	✓				Improved particle swarm optimization (IPSO)	Classification of mental task from EEG data
Bashir and El-Hawary [198]	ANN	✓				Particle swarm optimization (PSO)	Modeling hourly load forecasting problem

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Kiranyaz et al. [199]	FFNN	✓		✓		Particle swarm optimization (PSO)	Synthetic problems
Khayat et al. [133]	SOFNN	✓				Particle swarm optimization (PSO) & GA	Three tested examples
Tong and Mintram [21]	FFNN	✓				Genetic algorithm (GA)	Real-world applications (feature selection)
Slowik [200]	FFNN	✓				Differential evolution (DE) algorithm	Continuous optimization problems
Kordík et al. [201]	FFNN	✓				Meta-heuristic algorithms (MH)	Several real-world problems and benchmark data sets
Lian et al. [202]	ANN	✓		✓		Particle swarm optimization (PSO)	Non-linear system identification
Cruz-Ramírez et al. [135]	ANN	✓	✓			Memetic pareto differential evolution (MPDE)	Growth multi-classes in predictive microbiology
Zhao et al. [203]	RBFNN	✓		✓		Particle swarm optimization (PSO)	Melt Index modeling and Prediction problems
Subudhi and Jena [29]	MLPNN	✓				Memetic differential evolution (MDE)	Nonlinear system identification
Ma et al. [204]	ANN	✓				Genetic algorithm (GA)	Modeling chemical oxygen demand removal
Ding et al. [205]	FFNN	✓				Genetic algorithm (GA)	Real-world applications (The UCI data set)
Subudhi and Jena [206]	FFNN	✓				Opposition based differential evolution (ODE)	Nonlinear system identification

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Ghalambaz et al. [207]	MLPNN	✓				Gravitational search algorithm (GSA)	Wessinger's Equation
Irani and Nasimi [208]	FFNN	✓				Genetic algorithm (GA)	Permeability estimation of the reservoir
Li and Liu [209]	RBFNN	✓				Modified PSO simulated annealing (MPSOSA)	Melt index prediction model
Sun et al. [210]	NN	✓				Genetic algorithm (GA)	Dynamic prediction of financial distress
Ozbakır and Delice [211]	MLPNN	✓✓				Binary particle swarm optimization (BPSO)	Exploring comprehensible classification rules
Carvalho et al. [212]	FFNN	✓				VNS, SA, GEO, and GA algorithms	Identification and estimation of pollution sources
Han et al. [213]	FFNN	✓				Gaussian particle swarm optimization (GPSO)	Predictive control and system identification
Zhao and Qian [137]	3LFFNN	✓		✓		Cooperative particle swarm optimization (CPSO)	The application of predicting the sunspot numbers
Zanchettin et al. [214]	MLPNN	✓	✓	✓		Simulated annealing (SA), Tabu search (TS) and GA	Data classification
Vadood et al. [215]	ANN		✓	✓	✓	Genetic algorithm (GA)	Optimization of acrylic dry spinning production line
Malviya and Pratihar [136]	RBFNN	✓			✓	Particle swarm optimization (PSO)	Metal inert gas (MIG) welding process

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Vasumathi and Moorthi [139]	ANN	✓				Particle swarm optimization (PSO)	power engineering optimization problem
Mirjalili et al. [216]	FFNN	✓				Hybrid PSO &gravitational search algorithm (GSA)	Three benchmark problems
Khan and Sahai [217]	FFNN	✓				Bat algorithm (BA), GA & PSO	Standard dataset (in the field of Medicine)
Huang et al. [218]	RBF	✓				Improved chaos optimization (ICO)	Melt index prediction
Green II et al. [138]	FFNN	✓				Central force optimization (CFO) & PSO	Data classification
Irani and Nasimi [219]	BPNN	✓				Ant colony optimization (ACO)	Permeability Estimation of the Reservoir
Kulluk et al. [220]	FFNN	✓				Self-adaptive global best harmony search (SGHS)	six benchmark classification problems
Nandy et al. [221]	FFNN	✓				Firefly optimization algorithm (foa)	Iras dataset, Wine dataset and Liver dataset
Yaghini et al. [140]	ANN	✓				Improved particle swarm optimization (IPSO)	Eight benchmark datasets
Han and Zhu [222]	FFNN	✓				Improved particle swarm optimization (IPSO)	Function approximation and classification problems
Sharma et al. [223]	FFNN	✓				Ant colony optimization (ACO) algorithms	Bankruptcy prediction in banks

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Li et al. [224]	GRNN	✓		✓		Fruit fly optimization algorithm (FOA)	Annual power load forecasting
Ismail et al. [142]	PUNN	✓				Particle swarm optimization (PSO)	Load–deformation analysis of axially loaded piles
Wang et al. [225]	ANN	✓				Group search optimization (GSO)	Spatiotemporal prediction for nonlinear system
Lu et al. [226]	QNN	✓	✓			Quantum-based algorithm (QBA)	Several Benchmark Classification problem
Askarzadeh and Rezazadeh [227]	FFANN	✓				Bird mating optimizer (BMO)	Three real-world classification problems
Li et al. [228]	FFNN	✓				Convex combination algorithm (CCA)	Several computational experiments
Dragoi et al. [141]	ANN	✓	✓		✓	Self-adaptive differential evolution algorithm (SADE)	An aerobic fermentation process
Parra et al. [229]	ANN	✓			✓	Evolutionary strategy (ES)	Time series, classification and biometric recognition
Mirjalili et al. [144]	MLPNN	✓	✓	✓		Biogeography-based optimization (BBO)	5 classification and 6 function approximation datasets
Piotrowski [230]	MLPNN	✓				Differential evolution (DE)	Real-world regression problem & Benchmark problems

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Nasimi and Irani [231]	ANN	✓				Particle swarm optimization (PSO)	Identification and modeling of a yeast fermentation bioreactor
Tapoglou et al. [232]	FFNN	✓				Particle swarm optimization (PSO)	Groundwater-level forecasting under climate change scenarios
Raja et al. [233]	DENN	✓				Particle swarm optimization (PSO)	Bratu equation arising in the fuel ignition model
Beheshti et al. [234]	MLPNN	✓				Centripetal accelerated PSO (CAPSO)	Medical diseases diagnosis
Ren et al. [235]	BPNN	✓				Particle swarm optimization (PSO)	Wind speed forecasting (WSF) problem
Das et al. [143]	ANN	✓	✓	✓	✓	Particle swarm optimization (PSO)	Non-linear channel equalization problem
Jaddi et al. [145]	ANN	✓	✓			Multi-population cooperative bat algorithm	Classification and time series prediction benchmark datasets
Svečko and Kusić [236]	FFNN	✓				BAT search algorithm	The precise positional controls of piezoelectric actuators
Kumaran and Ravi [237]	ANN	✓				Biogeography-based optimization (BBO)	Long-term sector-wise electrical energy forecasting
Cui et al. [238]	SMRNNN	✓				Improved glowworm swarm optimization (IGSO)	Time series prediction
Chen et al. [239]	NFN	✓			✓	Improved artificial bee colony (IABC)	Approximation of the Piecewise Function

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Mirjalili [240]	MLPNN	✓				Grey Wolf optimizer (GWO)	Five classification and three function-approximation DB
Agrawal and Bawane [241]	ANN			✓		Swarm optimization (PSO)	Pixel classification in satellite imagery
Gharghan et al. [242]	ANN			✓	✓	Particle swarm optimization (PSO)	Indoor and outdoor track cycling problem
Vadood et al. [243]	ANN		✓	✓		Genetic algorithm (GA)	Prediction of resilient modulus of polyester
González et al. [147]	NN		✓	✓		Fuzzy gravitational search algorithm (FGSA)	Particular pattern recognition application (medical images)
Jaddi et al. [146]	ANN	✓	✓	✓		Modified bat-inspired algorithm (MBA)	classifications and time series datasets
Gaxiola et al. [148]	T2FNN	✓				Particle swarm optimization (PSO) & genetic algorithm	Mackey–Glass time series problem
Razmjooy and Ramezani [30]	WNN	✓				Hybrid PSO & gravitational search algorithm	System identification
Yazdi et al. [244]	NN	✓				Artificial bee colony (ABC)	Optimization of geometrical parameters
Jia et al. [245]	RBFNN	✓		✓		Genetic algorithm (GA)	Classification of Small Samples (benchmark)

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Leema et al. [246]	FFANN	✓				Differential evolution (DE) & PSO	Three benchmark clinical datasets
Karaboga and Kaya [149]	ANFIS	✓				Hybrid artificial bee colony (aABC)	Nonlinear dynamic systems
Xia et al. [247]	RBFNN	✓				Bare-bones particle swarm optimization (BBPSO)	Starch foam material performance prediction
Melo and Watada [248]	FFNN	✓		✓		Gaussian-particle swarm optimization (GPSO)	The Iris data classification problem
Chidambaram et al. [249]	ANN		✓	✓	✓	Genetic algorithm (GA)	Prediction of the base plate temperature of the fin
Khishe et al. [19]	MLPNN	✓				Improved biogeography-based optimization (IBBO)	Sonar dataset classification
Pradeepkumar and Ravi [250]	QRNN	✓				Particle swarm optimization (PSO)	Forecasting Financial Time Series Volatility
Islam et al. [251]	ANN	✓				Chaotic genetic algorithm-simulated annealing (SA)	Electrical energy demand prediction in smart grid
Emary et al. [252]	FFNN	✓				Grey Wolf optimizer (GWO)	Feature Selection and classification problems
Taheri et al. [253]	ANN	✓				Hybrid artificial bee colony (HABC)	Forecasting the blast-produced ground vibration
Chatterjee et al. [254]	MLPFFNN	✓				Particle swarm optimization (PSO)	Structural failure prediction of multistoried RC buildings

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Song et al. [255]	DNN	✓				Particle swarm optimization (PSO)	Transient probabilistic analysis of flexible mechanism
Yan et al. [256]	BRNN	✓				Particle swarm optimization (PSO) algorithm	Stock prediction
Ganjefar and Tofighi [151]	QNN	✓				Hybrid genetic algorithm (HGA)	Function approximation problem
Jafrasteh and Fathianpour [150]	LLRBFNN	✓				Artificial bee colony (SPABC)	Ore grade estimation
Aljarah et al. [152]	MLPNN	✓				Whale optimization algorithm (WOA)	Benchmark datasets
Mansouri et al. [257]	ANN	✓				Grey Wolf optimizer (GWO)	Anomaly recognition in industrial sensor networks
Rukhaiyar et al. [258]	ANN	✓				Particle swarm optimization (PSO)	Predicting factor of safety of slope problem
Semero et al. [259]	FFNN	✓				Particle swarm optimization (PSO) & GA	Short-term wind power forecasting
Bohat and Arya [260]	FFNN	✓				Gbest-guided gravitational search algorithm (GSA)	Real-Parameter Optimization
Mostafaeipour et al. [261]	MLPNN	✓				BA & firefly optimization algorithm (FOA)	Prediction of air travel demand

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Camci et al. [262]	T2FNN	✓				PSO-sliding mode control (PSOSMC)	Agricultural robots, or agrobots
Hadavandi et al. [154]	MLPNN	✓				Grey wolf optimizer (GWO)	Modeling the strength of siro-spun yarn in spinning mills
Huang and Liu [263]	RBF	✓		✓	✓	Particle swarm optimization (PSO)	Price Forecasting Method of Carbon Trading Market
Nayak and Misra [264]	CPNN	✓				Genetic algorithm (GA)	The estimating stock closing indices problem
Agrawal et al. [265]	RBFNN	✓				Fuzzy particle swarm optimization (PSO)	Multi-label classification & real-world datasets
Mao et al. [266]	T2FNN	✓				Grey wolf optimizer (GWO)	Single input/output and multi-input/output systems
Tian et al. [267]	ANN			✓		Genetic algorithm (GA)	Detection of loss of nuclear power plants
Tang et al. [268]	FFANN	✓	✓	✓		Dynamic group optimisation (DGO)	Approximation testing function
Haznedar and Kalinli [155]	ANFIS				✓	Simulated annealing (SA)	Dynamic systems identification problems
Xu et al. [269]	FFANN	✓				Modified artificial bee colony (MABC)	Benchmark functions

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Heidari et al. [153]	MLPNN	✓				Grasshopper optimization algorithm (GOA)	Medical diagnosis classification datasets
Karkheiran et al. [270]	FFBPNN	✓				Particle swarm optimization (PSO) & GA	Precise estimation of the local scour at bridge piers
Ong and Zainuddin [271]	WNN			✓		Modified cuckoo search algorithm (MCS)	Multi-step ahead chaotic time series prediction
Harandizadeh et al. [272]	ANFIS	✓	✓			Particle swarm optimization (PSO)	Prediction of pile bearing capacity problem
Pham et al. [156]	MLPNN	✓			✓	Biogeography-based optimization (BBO)	Predicting coefficient of consolidation of soil
Han et al. [157]	FFNN	✓				Differential evolution (DE)	Prediction of cooling efficiency of forced-air systems
Jiang et al. [273]	BPNN	✓				Genetic algorithm (GA)	Power Grid Investment Risk (PGIR) problem
Xu et al. [274]	BPNN	✓				Grey wolf optimizer (GWO)	Prediction of mobile multiuser communication networks
Djema et al. [275]	MLPNN	✓				Grey wolf optimizer (GWO)	Adaptive direct power control problem

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Li et al. [276]	GRNN	✓				Cuckoo search algorithm (CS)	Power transformer fault diagnosis problem
Zhao et al. [277]	MLPNN	✓				Selfish herd optimization algorithm (SHO)	UCI machine learning repository
Faris et al. [278]	FFNN	✓		✓		Grey wolf optimizer (GWO)	Twenty-three standard classification datasets
Rojas-Delgado et al. [158]	ANN	✓				PSO & FOA & cuckoo search (CS)	Six classification benchmark datasets
Bui [279]	ANN	✓				BBO, GSA and GWO	Forest fire susceptibility mapping in Dak Nong
Yu and Zhao [280]	BPNN	✓				Genetic algorithm (GA)	Prediction of critical properties of biodiesel fuels
Ma et al. [281]	NCLNN	✓				Particle swarm optimization (PSO)	Forecasting short-term wind speed of wind farms in China
Wang et al. [160]	MLFFNN	✓				Human-behavior PSO & cellular automata (CA)	15 benchmark complex and real-world datasets
Son et al. [53]	NNARX	✓				Jaya algorithm (JA)	Uncertain nonlinear system identification

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Raval and Pandya [282]	NNFS	✓		✓		Particle swarm optimization (PSO)	Extra High Voltage Transmission lines
Kuntoji et al. [283]	ANN		✓	✓		Particle swarm optimization (PSO)	Prediction of wave transmission
Al-Majidi et al. [161]	FFNN	✓	✓			Particle swarm optimization (PSO)	Predicting the maximum power point of a photovoltaic array
da Silva Veloso et al. [284]	FFNN			✓	✓	Particle swarm optimization (PSO)	The spouted bed drying of deformable solid materials
Yadav and Satyannarayana [285]	FFNN	✓		✓	✓	Multi-objective genetic algorithm (MOGA)	Estimating suspended sediment yield
Wu et al. [286]	ANN	✓		✓	✓	Particle swarm optimization (PSO)	Prediction of Endpoint Sulfur Content in KR Desulfurization
Ertuğrul [54]	RaANN			✓	✓	Differential evolution algorithms (DE)	48 synthetic datasets
Khishe and Mosavi [159]	MLPNN	✓				Chimp optimization algorithm (ChOA)	Classification of underwater acoustical dataset
Shen et al. [287]	BPNN	✓				Particle swarm evolution (PSE)	Microchananel resistance factor prediction
Huang et al. [288]	BPNN	✓				Improved particle swarm optimization (IPSO)	Air Quality Prediction
Shen et al. [289]	BPNN	✓				Genetic algorithm (GA)	Forecasting Model for the Velocity of Robotic Fish

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Ghanem et al. [290]	BPNN	✓				ABC and dragonfly algorithm (DA)	Efficient Intrusion Detection Model
Ansari et al. [162]	BPNN	✓				Magnetic optimization algorithm (MOA) & PSO	Bankruptcy Prediction problem
Gong et al. [291]	ANN	✓				Whale optimization algorithm (WOA)	Brain tumor diagnosis
Zeng et al. [292]	ANN	✓				Fruit fly optimization algorithm (FOA)	User equipment association in wireless sensor
Supraja et al. [293]	ANN	✓				GA & Shuffled frog-leaping algorithm (SFLA)	Prediction of free spectrum in cognitive radio
Fang et al. [294]	MLPNN	✓				Whale optimization algorithm (WOA)	Automatic breast cancer detection
Zafar et al. [295]	ANN	✓				Particle swarm optimization (PSO)	Internet of Things (IOT)
Darabi et al. [296]	ANN	✓				Grey Wolf optimizer (GWO)	Spatial prediction of urban flood-inundation
Qiao et al. [297]	MLPNN	✓				Whale optimization algorithm (WOA)	Underwater targets classification
Zheng et al. [298]	FFNN	✓				Salp swarm optimization (SalpSO)	Resources Policy
Bahiraei et al. [299]	ANN	✓				Ant lion optimizer (ALO) algorithm	Predicting heat transfer rate
Zhang et al. [163]	Elman NN	✓	✓	✓		Chicken swarm optimization (CSO)	Air pollution

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Njock et al. [300]	ANN		✓	✓	✓	Differential evolution (DE)	Mechanics and Geotechnical Engineering
Khatir et al. [301]	ANN	✓				Arithmetic optimization algorithm (AOA)	Damage assessment in FGM composite plates
Yeganeh and Shadman [302]	ANN	✓	✓	✓		GA & PSO	Monitoring binary and polytomous logistic profiles
Guo et al. [303]	RBFNN				✓	JAYA optimization algorithm	Energy storage systems problems
Korouzhdeh et al. [304]	ANN			✓	✓	Biogeography-based optimization (BBO)	Construction and Building Materials
Li et al. [305]	RBFNN	✓				Fruit fly optimization algorithm (FOA)	Vegetable price forecasting
Cui et al. [306]	BPNN	✓				Biogeography-based optimization (BBO)	Multiple-criteria inventory classification
Bai et al. [307]	BPNN	✓				Improved particle swarm optimization (PSO)	Reliability prediction in engineering
Ghersi et al. [308]	ANN	✓				Genetic algorithm (GA)	Optimization of power and generation engines by biogas
Luo et al. [309]	FFNN	✓				Spotted hyena optimizer (SHO)	Three function-approximations
Fetimi et al. [310]	ANN	✓				Particle swarm optimization (PSO)	Environmental Chemical Engineering
Yibre and Koçer [311]	FFNN	✓				Artificial algae algorithm (AAA)	Semen quality predictive model

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Sun et al. [312]	Elman NN	✓				Quantum water strider algorithm (QWSA)	Energy estimation
Sheelwant et al. [313]	ANN	✓				Genetic algorithm (GA)	Communications (aluminum metal matrix composites)
Medi and Asadbeigi [314]	NNARX	✓				Genetic algorithm (GA)	Nonlinear chemical and biochemical processes
Zhang et al. [315]	BPNN	✓				Chaotic adaptive gravity search algorithm (CAGSA)	Fault diagnosis of electrical machine drive system
Zhao et al. [316]	BPNN	✓				Whale optimization algorithm (WOA)	Prediction of the deflection of reinforced concrete beams
Garcıa-Rodenas et al. [317]	FFNN	✓				Memetic chaotic gravitational search algorithm (MCGSA)	Approximation of a continuous function
Uzlu [318]	ANN	✓				Grey wolf optimizer (GWO)	Estimates of greenhouse gas emission
Saffari et al. [319]	MLPNN	✓				Chimp optimization algorithm (ChOA)	Marine mammal classification
Liu et al. [320]	FNN	✓				Particle swarm optimization (PSO)	Path planning problem
Bui et al. [321]	ANN	✓				Cuckoo search optimization (CSO)	Predicting Ground Vibrations
Raei et al. [322]	BPNN	✓				Whale optimization algorithm (WOA)	Soil wind erodibility
Cui et al. [323]	BPNN	✓				Genetic algorithm (GA)	Applications in prediction of foundation pit deformation

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Sağ and Abdullah Jalil [324]	FFNN	✓				Vortex search (VS) Optimization algorithm	Classification Dataset
Wang et al. [325]	ANN	✓				Genetic algorithm (GA)	Prediction of parameters of shot peen forming
Wang et al. [326]	BPNN	✓				Whale optimization algorithm (WOA)	Image denoising
Turki and Shammari [327]	FFNN			✓		Genetic algorithm (GA)	Predicting the Output Power of a Photovoltaic Module
Eappen et al. [328]	ANN	✓				Advanced squirrel algorithm (ASA)	Cognitive radio-based air traffic control application
BACANIN et al. [329]	ANN	✓		✓		Artificial bee colony (ABC)	Five well-known medical benchmark datasets
Liu et al. [330]	BPNN	✓				Hybrid GA-PSO	Data fusion for multi-source sensors
Nguyen et al. [331]	BPNN	✓				Accelerated particle swarm optimization (APSO)	Robot precision positioning
Ge et al. [332]	Regression NN				✓	Grey wolf optimizer (GWO)	Short-term load forecasting of regional distribution network
Kaur and Chahal [333]	ANFIS				✓	Particle swarm optimization (PSO)	Prediction of Chikungunya disease

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Zhang et al. [334]	BPNN	✓				Improved grey wolf optimizer (IGWO)	Energy Storage
Guo et al. [335]	ELMAN NN	✓				Whale optimization algorithm (WOA)	Monophenolase assay-analytical biochemistry
Xue et al. [336]	FFNN	✓				Differential evolution (DE)	Different classification problems
Ding et al. [337]	ANN	✓				Jaya algorithm (JA)	Simultaneous identification of structural damage
Zhu et al. [338]	ANN	✓				Adaptive genetic algorithm (AGA)	Wave energy converter arrays
Jnr et al. [339]	BPNN	✓				Aquila optimization algorithm (AOA)	Wind speed prediction
Zhao et al. [340]	ANN	✓				Multi-tracker optimization algorithm (MTOA)	Predicting compressive strength of concrete
Wua et al. [341]	ANN	✓				Bees algorithm (BA)	Welding sequence Engineering optimization
Si et al. [342]	MLPNN	✓				Equilibrium optimizer (EO) algorithm	Medical data classification
Khan et al. [343]	FLNN	✓				Accelerated particle swarm optimization (APSO)	Medical data classification
Li et al. [164]	MLPNN	✓				Biogeography-based optimization (BBO)	Medical data classification

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Gülcü [344]	MLPNN	✓				Dragonfly algorithm (DA)	Real-world civil engineering and classification datasets
Netsanet et al. [345]	ANN	✓				Ant colony optimization (ACO)	Short-term PV power forecasting
Liang et al. [346]	MLPNN	✓				Hunger games search optimization (HGSO)	Building Engineering
Chondrodima et al. [347]	RBFNN	✓				Particle swarm optimization (PSO)	Public transport arrival time prediction
Ehteram et al. [348]	MLPNN	✓				Multi-objective salp swarm algorithm (MOSSA)	Predicting evaporation
Li et al. [349]	Elman NN	✓				Sparrow search algorithm (SSA)	Thermal error modeling of motorized spindle
Ibad et al. [350]	Spiking NN				✓	Salp swarm algorithm (SSA)	Time-Series Classification Problem
Foong and Moayedi [351]	MLPNN	✓				Equilibrium optimization (EO) & VSA	Slope stability evaluation
Chatterjee et al. [352]	FFNN	✓				Chaotic whale optimization algorithm (COWOA)	Classification dataset
He et al. [353]	CFNN	✓				Grey wolf optimizer (GWO)	Predicting the compressibility of clay
Gülcü [354]	MLPNN	✓				Improved animal migration optimization (IAMO)	Classification dataset

Table 2 (continued)

Authors & dates	Neural network categories	Optimized components: 1. Weights & bias, 2. Layers 3. Nodes 4. Activation function and learning parameters				The meta-heuristic algorithm used for training neural networks	Application / dataset
		1	2	3	4		
Liu et al. [355]	BPNN	✓				Genetic algorithm (GA)	Electrical Engineering & Technology
Bataineh et al. [356]	MLPNN	✓				Clonal selection algorithms (CSA)	Five classification datasets
Han et al. [357]	FNN	✓			✓	Multi-objective PSO (MOPSO)	Nonlinear Systems Identification
Deepika and Balaji [358]	ANN	✓			✓	Differential evolution (DE)	Effective heart disease prediction problem
Kirankaya and Aykut [359]	ANN	✓				Artificial bee colony (ABC) algorithms	Classification dataset
Yan et al. [360]	MLPNN	✓				Chaotic grey wolf optimization (CGWO)	Energy
Li et al. [361]	BPNN	✓				Genetic algorithm (GA)	Coastal Bulk (Coal) Freight Index Forecasting
Kuo et al. [362]	BPNN	✓				Simulated annealing (SA)	Classification dataset (MNIST and FASHION)
Zhao et al. [363]	BPNN	✓				Sparrow search algorithm (SSA)	Predicting the Thickness of an Excavation Damaged Zone
Davar et al. [364]	BPNN	✓				Butterfly optimization algorithm (BOA) & PSO	Predicting Matric Suction in Expansive Clay Soil
Huang et al. [365]	BPNN	✓				Firefly algorithm (FA)	Micromachined Silicon Resonant Accelerometers
Wang et al. [366]	RBFNN	✓				Grey wolf optimizer (GWO)	Electrical Impedance Tomography

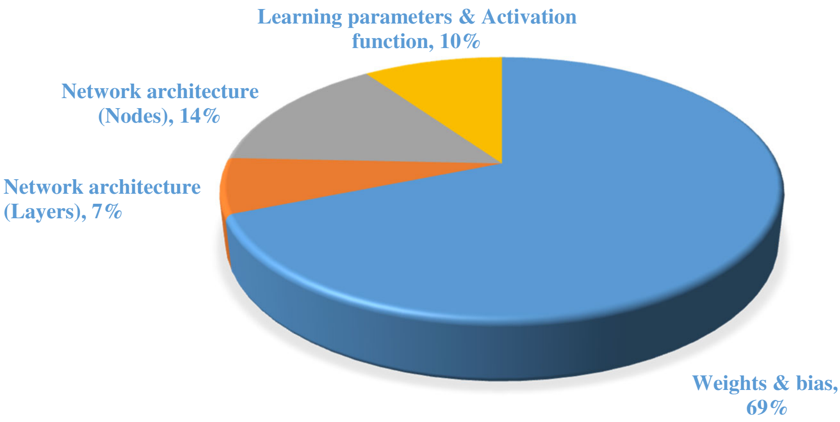

Fig. 12 Relative abundance of research on optimized components in ANNs using MH algorithms

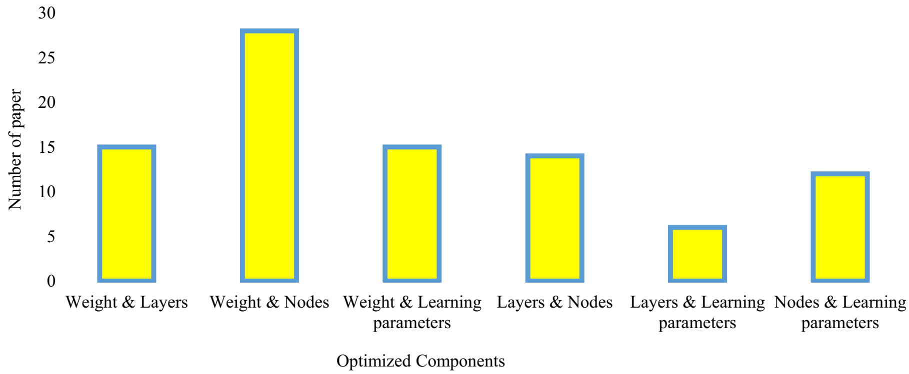

Fig. 13 Relative abundance of research in the simultaneous optimization of two components of ANNs using MHs

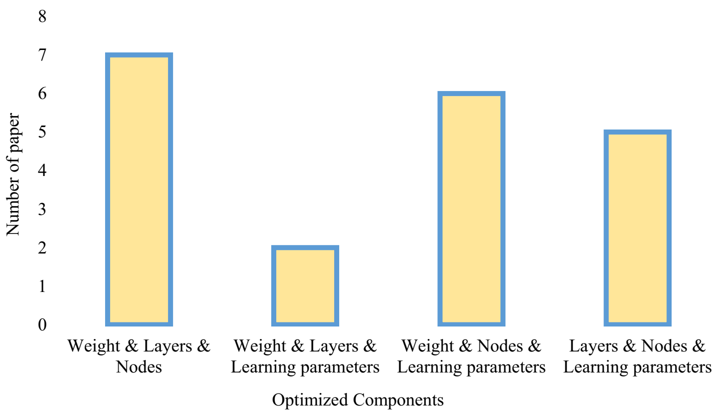

Fig. 14 Relative abundance of research in the simultaneous optimization of three components of ANNs using MHs

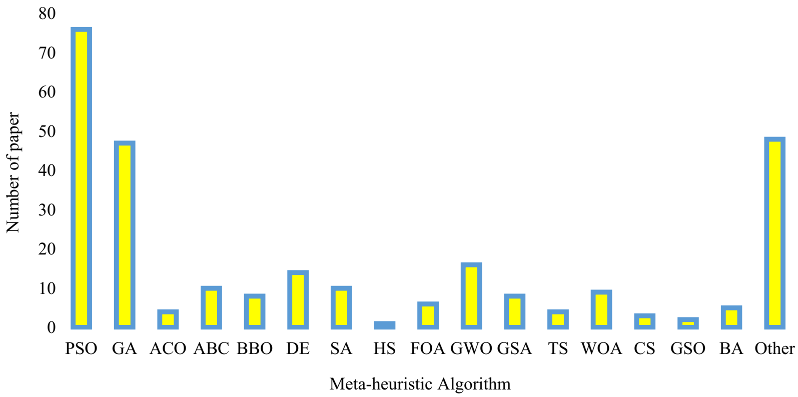

Fig. 15 Meta-heuristic algorithms used to optimize ANNs

4.1.2 Investigation of Meta-Heuristic Algorithms Used in Ann's Optimization

According to Table 2, many MH algorithms have been developed to optimize neural networks. Figure 15 shows the MH algorithms used to optimize ANNs. PSO, 76 implementations and GA, 47 implementations, was the most used MH algorithms. GWO, DE, SA, ABC, GSA, WOA, BBO, and FOA algorithms are also in the next ranks. Most researchers tend to extend novel hybrid algorithms by combining MHs to optimize the hyper-parameters of ANNs. The development of hybrid MHs helps improving algorithms performance and capable of solving complex optimization problems. According to the results of Table 2, many researches have used the modification and hybridization of meta-heuristic algorithms to optimize neural network parameters. Also, the performance of the proposed hybrid MH algorithms have been better than others.

4.1.3 Checking the Number of Papers Published in Journals and Years

In this section, the papers in Table 2 are categorized according to the type of journals and the year of their publication. Figure 16 shows the percentage of papers published in various journals (based on Table 2). As shown, 74 papers (44%) in Elsevier, 30 papers (21%) in Springer, 27 papers (13%) in IEEE, 16 papers (8%) in Taylor & Francis, 13 papers (6%) in John Wiley & Sons, and 14 papers (8%) in other journals have been published regarding the use of MH for ANNs.

Figure 17 also indicates the changes in the number of papers published in different years about the use of MH for Training ANNs. Between 1988 and 2002, few papers were developed for neural network optimization. From 2003 to 2010, neural network optimization received a little more attention from researchers, and the number of papers in this field increased. But from 2011 to 2022, many researchers have worked on neural network optimization. Especially since 2021, the number of these papers has been increasing. This implies that this problem is still a challenge and many problems need to be resolved.

4.1.4 Applications of Hybrid MH-NNs

In this section, the application of the papers in Table 2 is evaluated. Figure 18 shows the

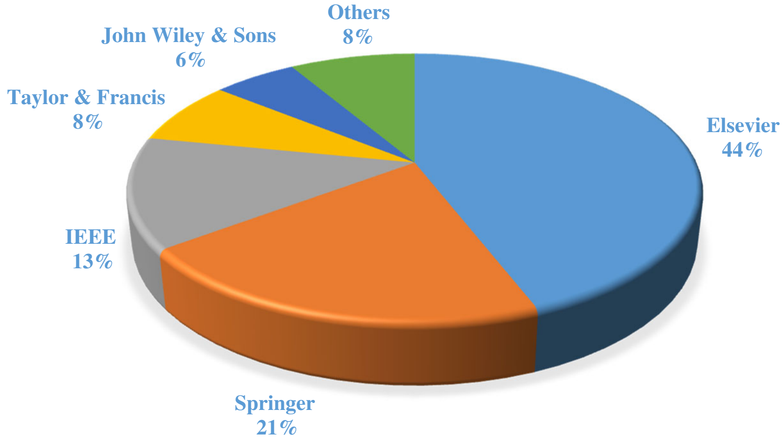

Fig. 16 Papers published in journals (based on Table 2)

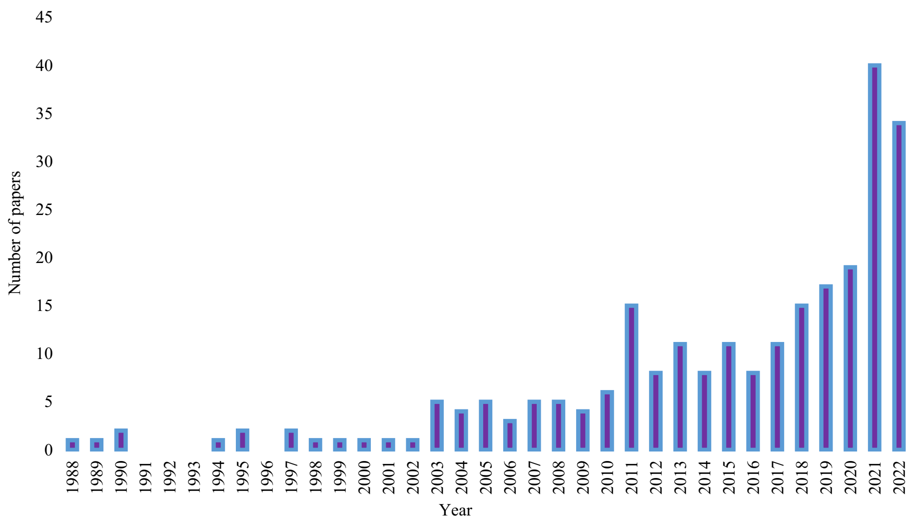

Fig. 17 Changes in the number of papers published in different years about the use of MH for Training ANNs

application of the papers regarding the use of MH for ANNs. 77 papers in benchmark problem (Classification, prediction, time series, optimization, system identification), 53 papers in electrical engineering, signal processing and energy systems, 34 papers in civil engineering, 18 papers in mechanical engineering, 16 papers in biomedical and chemical engineering, 15 papers in medical image classification and medical diseases diagnosis, 8 papers in environmental management, 8 papers in economy and product quality, and 19 papers in other applications have been published regarding the use of MH for ANNs.

As can be seen, most of the MH-ANNs were implemented on benchmark problems and datasets. The optimal solutions of the benchmark problems are known. Therefore, they are a very good criterion for evaluating algorithms. Also, many evolutionary ANNs have been

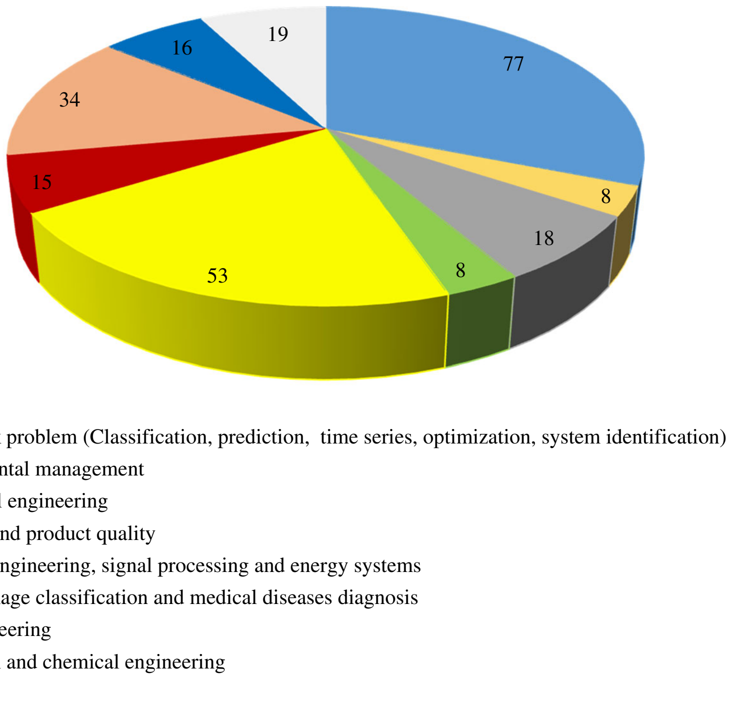

Fig. 18 Application of papers regarding the use of MH for ANNs

implemented in electrical engineering, civil engineering, mechanical engineering, and medical image classification applications. The results of these papers show that the proposed hybrid ANNs architectures perform better than others. Therefore, it can be said that evolutionary artificial neural networks (MH-ANNs) are promising methods in these applications.

4.1.5 Contributions of Different Continents in Using the Hybrid MH-NN Models

Figure 19 shows the distribution of studied papers according to the affiliation of the authors for each continent. As can be seen, Asia has the largest portion of contributions in the world with the maximum number of papers from China, Korea, and India, while America has the lowest contributions.

4.2 Review2: Training the DL Architectures by MH Algorithms

One of the weaknesses of DL architectures is finding the optimal value of algorithm parameters. This section provides a comprehensive overview of optimizing different DL architectures using MH algorithms. Optimization of DL architectures is often considered from several aspects: optimization of weights, hyper-parameters, network structure, activation nodes, learning parameters, learning algorithm, learning environment, etc. [9].

Ku et al. [367] used the genetic algorithm to optimize the weights of an RNN. The proposed approach (GA-RNN) was compared with Lamarckian and Baldwinian mechanisms, which indicated better results (convergence speed and accuracy). Blanco et al. [368] used the

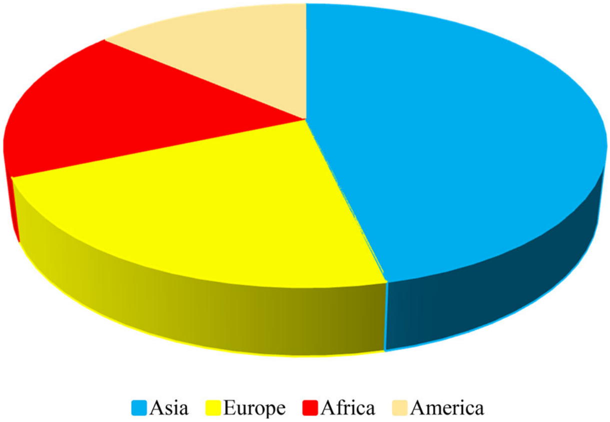

Fig. 19 Contributions of different continents in using the hybrid MH-NN models

genetic algorithm (GA) to improve the performance of an RNN. The results indicated that the proposed algorithm solves the time complexity well. Delgado et al. [369] used multi-objective SPEA2 and NSGA_II algorithms to optimize the topology and structure of an RNN. The proposed architectures performed well for the time series problem. Bayer et al. [370] used the NSGA_II to train an LSTM architecture. The results showed that the proposed network performs well in learning sequences.

Lin and Lee [371] used the improved PSO algorithm to optimize the weights of an RFNN. The results indicated that the IPSO algorithm for controlling nonlinear systems performed better than other methods (traditional PSO and GA). Subrahmanya and Shin [372] used the combination of PSO and CMA-ES algorithms to optimize the structure and weights of an RNN. According to the results, the proposed architecture (HMH-RNN) indicated good performance. Hsieh et al. [373] used the artificial bee colony (ABC) algorithm to optimize the weights of an RNN. According to experiments, the proposed approach indicates good capital market performance and can be implemented in a trading system to predict stock prices and maximize profits.

David and Greental [41] used combined gradient-based learning and genetic algorithm strategy to train a deep neural network. The proposed architecture performed very well in the benchmark data set. Shinozaki and Watanabe [40] used GA and CMA-ES algorithms to optimize the structure and parameters of a DNN. The results demonstrated that the proposed algorithm is suitable for adjusting neural network parameters. Sheikhan et al. [374] used the GSA binary algorithm to optimize the structure and weights of an RNN network. The proposed algorithm (BGSA-RNN) was compared with gradient-based and PSO algorithms, which provided significant results. A combination of evolutionary algorithm and DBN network was used by Chen et al. [375] for image classification. The results indicated that the execution time decreases rapidly.

Real et al. [376] used an evolutionary algorithm for convolutional neural network (CNN) training to classify CIFAR-10 and CIFAR-100 datasets. The findings implied that the proposed approach could provide competitive results in two popular datasets. Tang et al. [377] used the PSO algorithm to optimize the weights of a DSNN. The proposed algorithm performed very well in feature extraction problems and EEG signal detection. Song et al. [378] used improved biogeography-based optimization (IBBO) to optimize the parameters and weights of DDEA. The results indicated that the proposed approach (IBBO-DDEA) for gastrointestinal complications prediction performed better than other methods (such as ANN and other common architectures).

Da Silva et al. [379] used the PSO algorithm to optimize the hyper-parameters of a convolutional neural network. Experiments on a CAD system indicated an improvement in the accuracy of the proposed algorithm. The WWO algorithm was used by Zhou et al. [380] to optimize the structure and weights of a DNN. Experiments on several benchmark datasets indicated that the proposed WWO-DNN approach performs better than the gradient-based methods. Shi et al. [381] used the PSO algorithm to optimize the number of neurons in the hidden layers of a deep neural network. Experimental results demonstrated that the detection rate in the proposed algorithm was improved by 9.4% and 8.8% compared to conventional DNN and support vector machine (SVM). In addition, another experiment compared to the genetic algorithm (GA) proved that the proposed particle swarm optimization (PSO) is more effective in deep neural network (DNN) optimization. Hong et al. [382] used the genetic algorithm (GA) to optimize the parameters and hyper-parameters of the CNN. Experimental results for the price forecasting problem showed that the proposed GA-CNN always offers higher forecasting accuracy and lower error rates than other forecasting methods.

Guo et al. [383] used a distributed particle swarm optimization (DPSO) algorithm to optimize the hyper-parameters of convolutional neural network (CNN). Experiments on the image classification dataset indicated that the proposed DPSO method improved the performance of the CNN model while reducing computational time compared to traditional algorithms. ZahediNasab and Mohseni [384] used the genetic algorithm (GA) to optimize the deep neural network (DNN) activation function. Experiments on the medical classification and MNIST datasets showed the proposed approach's superiority. It was also stated that selecting an appropriate adaptive activation function plays an important role in the quality of a deep neural network. Jallal et al. [385] used an improved PSO algorithm for DNN training to improve the prediction accuracy of a solar tracker. The DNN-RODDPSO algorithm performed better than the standard algorithms in the literature. Elmasry et al. [386] used the PSO algorithm to optimize the hyper-parameters of three DL algorithms called DNN, LSTM-RNN and DBN. Experiments on the network intrusion detection problem proposed that these three developed architectures performed better than conventional architectures.

Kan et al. [387] used the adaptive particle swarm optimization (APSO) algorithm to optimize the weights and biases of the convolutional neural network (CNN). According to the results, the proposed hybrid approach (APSO-CNN) performed well in IoT network intrusion detection. Also, the performance of the proposed hybrid architecture has been better than other algorithms. Kanna and Santhi, [388] used the black widow optimization (BWO) algorithm to optimize the weights of CNN-LSTM for intrusion detection systems. The results showed that the proposed hybrid architecture (BWO-CNN-LSTM) performs better than the other original architectures. Ragab et al. [389] used enhanced gravitational search optimization (EGSO) algorithm to optimize the weights and biases of the convolutional neural network (CNN). According to the results, the proposed hybrid approach (EGSO-CNN) performed well in COVID-19 diagnosis problem. Also, the performance of the proposed hybrid architecture has been better than other algorithms.

Table 3 summarizes the above research as well as many other studies. As can be seen, for each research, the author name, year of publication, type of DL, optimized components, type of MH algorithm used, application and data set used are listed. In the following, for a more comprehensive review, some statistical analysis of the research collected in Table 3 is presented.

Table 3 A summary of meta-heuristic algorithm developments for training/optimization deep learning architectures

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Ku et al. [367]	RNN	✓				Genetic algorithm (GA)	Prediction and classification problems
Blanco et al. [368]	RNN	✓				Real-coded genetic algorithm (GA)	Benchmark datasets
Delgado et al. [369]	RNN	✓	✓			Strength pareto evolutionary algorithm2 & NSGA_II	Time-series benchmark problem
Bayer et al. [370]	LSTM		✓			Non-dominated sorting genetic algorithm (NSGA-II)	Sequence learning
Subrahmanya and Shin [372]	RNN	✓	✓			PSO and CMA-ES	Tow MIMO non-linear processes
Lin and Lee [371]	RFNN	✓				Improved particle swarm optimization (IPSO)	Non-linear system control
Hsieh et al. [373]	RNN	✓				Artificial bee colony algorithm (ABC)	Several international stock markets
Cheung and Sable [390]	CNN		✓			Evolutionary algorithm (EA)	MNIST Variations, rectangles-image and image classification
David and Greental [41]	DNN				✓	Genetic algorithm (GA)	MNIST hand-written digit recognition database
Shinozaki and Watanabe [40]	DNN		✓		✓	Genetic algorithm (GA) & CMA-ES	Phoneme recognition and spoken digit detection tasks
Lander and Shang [42]	DAE	✓	✓	✓		Evolutionary algorithm (EA)	MNIST handwritten digits 1 k dataset

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Sheikhan et al. [374]	RNN	✓	✓			Binary gravitational search algorithm (BGSA)	Emotion recognition and speech processing
Desell et al. [391]	RNN	✓				Ant colony optimization (ACO)	Predicting general aviation flight data
Rosa et al. [43]	CNN			✓		Harmony search algorithm (HS)	Fingerprint and handwritten digit recognition
Chen et al. [375]	DBN				✓	Evolutionary function array classification voter (EFACV)	MNIST dataset
Rosa et al. [44]	DBN		✓		✓	Firefly algorithm (FA)	MNIST and Semeion Handwritten Digit datasets
Papa et al. [392]	DBN		✓		✓	Harmony search algorithm (HSA)	Binary image reconstruction
Zhang et al. [393]	DBN	✓	✓		✓	Multi-objective evolutionary algorithm (MOEA)	Remaining Useful Life Estimation in Prognostics
Tang et al. [377]	DSNN	✓				Particle swarm optimization (PSO)	Recognition of motor imagery EEG signals
Khalifa et al. [32]	CNN	✓				Particle swarm optimization (PSO)	Classification problem
Badem et al. [394]	DNN	✓				Hybrid artificial bee colony (HABC)	15 benchmark data sets

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Gelly and Gauvain [395]	RNN	✓				Particle swarm optimization (PSO)	Speech activity detection
Liu et al. [396]	CNN	✓	✓			Multi-objective evolutionary algorithm (MOEA)	The MNIST data set and the CIFAR-10 data set
Song et al. [378]	DDAE	✓	✓			Ecogeography-based optimization (EBO)	Predicting morbidity of gastrointestinal infections
ElSaid et al. [397]	LSTM-RNN		✓			Ant colony optimization (ACO)	Turbine engine vibration
Real et al. [376]	CNN	✓		✓		Evolutionary algorithm (EA)	The CIFAR-10 and CIFAR-100 datasets
Jiang et al. [22]	DNN		✓		✓	Modified genetic algorithm (MGA)	Demand Forecasting in Outpatient Department
Lopez-Rincon et al. [33]	CNN	✓		✓		Evolutionary algorithm (EA)	Cancer miRNA biomarkers classification
Ye [37]	DNN		✓	✓	✓	Particle swarm optimization (PSO)	Biological activity prediction datasets
Kim et al. [398]	DBN		✓			Particle swarm optimization (PSO)	Highly class imbalance problem
Fujino et al. [399]	CNN			✓		Genetic algorithm (GA)	Recognition of human sketches problem
Lorenzo et al. [400]	DNN			✓		Particle swarm optimization (PSO)	MNIST and CIFAR-10 dataset

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Dufourq and Bassett [34]	EDEN			✓		Genetic algorithm (GA)	Seven image and sentiment classification datasets
da Silva et al. [379]	CNN			✓		Particle swarm optimization (PSO)	Lung nodule false positive reduction on CT images
Chen et al. [401]	LSTM			✓		Extremal optimization algorithm (EO)	Wind speed forecasting
Passos et al. [402]	DBM			✓		Particle swarm optimization (PSO), AIWPSO, HS & IHS	Binary image reconstruction
Soon et al. [403]	CNN			✓		Particle swarm optimization (PSO)	Vehicle logo recognition
Peng et al. [38]	LSTM			✓		Evolutionary algorithm (EA)	Electricity price prediction problem
ElSaid et al. [404]	LSTM-RNN	✓	✓			Ant colony optimization (ACO)	Predict turbine engine vibration
Lorenzo and Nalepa [405]	DNN		✓			Memetic evolution algorithm (MEA)	segmenting medical images and CIFAR-10 benchmark
Pawełczyk et al. [406]	CNN	✓	✓			Genetic algorithm (GA)	MNIST set which contains grayscale images
Fielding and Zhang [407]	CNN			✓	✓	Particle swarm optimization (PSO)	CIFAR-10 image classification task
Martín et al. [45]	DNN		✓	✓	✓	Evolutionary algorithm (EA)	Dataset of handwritten digits images

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Sun et al. [408]	DNN	✓			✓	Evolutionary algorithm (EA)	Learning Meaningful Representations
Liang et al. [409]	DNN		✓	✓		Evolutionary algorithm (EA)	Omniglot Character Recognition problem
Wang et al. [35]	CNN		✓	✓		Hybrid differential evolution approach (HDE)	Image classification
Zhou et al. [380]	DNN	✓	✓			Water wave optimization (WWO)	Complex network optimization problems
Khodabandehlou and Fadali [410]	RNN	✓				Dynamical trajectory-based optimization (DTBO)	Three non- linear dynamical systems
Banharnsakun [46]	CNN	✓				Artificial bee colony (ABC)	MNIST handwritten image dataset (classification)
Gao and Li [411]	CNN	✓		✓		Segmented particle swarm optimization (SPSO)	Land cover and land use classification of RS images
Wang et al. [39]	CNN			✓		Particle swarm optimization (PSO)	Linear prediction model
Wang et al. [36]	GAN			✓		Evolutionary algorithm (EA)	Several image classification datasets
Fujino et al. [412]	CNN			✓		Genetic algorithm (GA)	Anime storyboard recognition problem
Li et al. [413]	CNN			✓		Particle swarm optimization (PSO)	Image classification
Li et al. [414]	DBNN		✓		✓	Multi-objective PSO (MOPSO)	Traffic flow forecasting

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Nepomuceno [415]	RRNN		✓			Multi-objective optimization (MOO)	System identification and modelling
Wei et al. [416]	DBN		✓			Artificial fish swarm algorithm (AFSA)-GA-PSO	Intrusion detection classification model
Shi et al. [381]	DNN		✓			Particle swarm optimization (PSO)	Digital modulation recognition
Junior and Yen [417]	CNN		✓	✓		Particle swarm optimization (PSO)	Image classification
Navaneeth and Suchetha [418]	1-D CNN				✓	Particle swarm optimization (PSO)	Real-time detection and classification applications
ZahediNasab and Mohseni [384]	CNN				✓	Genetic algorithm (GA)	CT brain and the MNIST hand written digits dataset
Goel et al. [419]	CNN			✓		Grey wolf optimizer (GWO)	An automatic diagnosis of COVID-19
Gao et al. [420]	CNN			✓		Gradient-priority particle swarm optimization (GPSO)	EEG-based Emotion Recognition
Martín et al. [421]	CNN	✓		✓		Hybrid statistically coral reef optimization (HSCRO)	The CIFAR-10 and the CINIC-10 datasets
Lan et al. [51]	CNN	✓				Particle swarm optimization (PSO)	Enhancing heart disease and breast cancer detection
Tang et al. [422]	LSTM	✓				Genetic algorithm (GA)	Traffic Flow Prediction on Urban Road Network

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Elmasry et al. [386]	LSTM-RNN			✓		Particle swarm optimization (PSO)	Network intrusion detection
Guo et al. [383]	CNN			✓		Distributed particle swarm optimization (DPSO)	Image classification benchmarks
Lima et al. [423]	CNN			✓		Simulating annealing (SA)	Toward classifying small lung nodules
Renukadevi and Karunakaran [424]	DBN	✓		✓		Grasshopper optimization algorithm (GOA)	Liver disease classification
Jallal et al. [385]	DNN		✓			Randomly occurring distributed delayed PSO	Monitoring the energy produced by solar trackers
Ali et al. [425]	DBN		✓	✓		Stacked genetic algorithm (SGA)	Heart Disease Prediction
Hong et al. [382]	CNN		✓	✓		Genetic algorithm (GA)	Locational Marginal Price Forecasting
Rajagopal et al. [426]	CNN		✓	✓		Multi-objective PSO (MOPSO)	Scene Classification in Unmanned Aerial Vehicles
Lu et al. [427]	CNN		✓	✓		Multi-objective genetic algorithm (MOGA)	Image Classification
Lin et al. [428]	DAE	✓				Ecogeography-based optimization (EBO)	In-Vehicle Networks-CAN Bus
Kavousi-Fard et al. [429]	GAN			✓		Modified firefly algorithm (MFA)	Securing Vehicles problem

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Johnson et al. [430]	CNN			✓		Genetic algorithm (GA)	Image classification dataset: CIFAR10, MNIST and Caltech
Kan et al. [387]	CNN	✓				Adaptive particle swarm optimization (APSO)	IoT network intrusion detection
Zheng et al. [431]	CNN			✓		Genetic algorithm (GA)	Pattern Recognition (parametric eye modeling)
Pang et al. [432]	CNN & LSTM			✓		Particle swarm optimization (PSO)	Hyperspectral imaging classification
Gai et al. [433]	DBN				✓	Sparrow search algorithm (SSA)	Detection of gear fault severity
Sun et al. [434]	DBN	✓				Improved archimedes optimization algorithm (IAOA)	Energy
Samir et al. [435]	CNN			✓		Heuristic-based JSO optimization algorithm	Predicting heart diseases problem
Liu et al. [436]	DNN			✓		Improved particle swarm optimization (IPSO)	COVID-19 spread
Maoa et al. [437]	CNN		✓	✓		Genetic algorithm (GA)	Waste classification—Image recognition
Gao et al. [420]	CNN			✓		Particle swarm optimization (PSO)	EEG-based emotion recognition

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Kim and Cho [438]	CNN-LSTM			✓		Particle swarm optimization (PSO)	Anomalous query access control
Zhang et al. [439]	CNN-LSTM			✓		Swarm-based optimization	Intelligent human action recognition
Li et al. [440]	CNN		✓			Sea lion insisted on dragon fly modification (SL-DU)	Hardening prediction in steel
Mohakud and Dash [441]	CNN			✓		Exponential grey wolf optimization (EN-GWO)	Skin cancer image segmentation
Martín et al. [421]	CNN	✓		✓		Hybrid coral reef optimization (HSCRO)	CIFAR-10 and the CINIC-10 Dataset
Altan et al. [442]	LSTM	✓				Grey wolf optimizer (GWO) Algorithm	Wind speed forecasting
Roder et al. [443]	DBN	✓				hill climb (HC) Metaheuristic optimization	Image classification Dataset
Mathe et al. [444]	CNN		✓	✓		Spider monkey-based electric fish optimization (SM-EFO)	Biomedical Signal Processing and Control
Mahesh et al. [445]	CNN			✓	✓	Jaya-based barnacle mating optimization (J-BMO)	Biomedical Signal Processing and Control
Singh et al. [446]	CNN			✓		Multi-level particle swarm optimization (MPSO)	Image classification Dataset
Kumar and Haider [447]	RNN–LSTM		✓	✓	✓	Flower pollination algorithm (FPA)	Prediction of Intra-day Stock Market

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Kumar et al. [448]	DNN		✓	✓	✓	Genetic algorithm (GA)	Four Image classification Dataset
Chitra and Kumar [449]	CNN		✓	✓	✓	Mutation-based atom search optimization (MASO)	Cervical cancer detection
Deighan et al. [450]	CNN			✓		Genetic algorithm (GA)	Gravitational wave classification
Qu et al. [451]	DAE			✓		Non-dominated sorting genetic algorithm (NSGA_II)	Classification problem
Goel et al. [452]	CNN	✓		✓		Grey wolf optimizer (GWO) algorithm	Spread of coronavirus disease (COVID-19)
Liu and Nie [453]	SSAE	✓				Invasive weed optimization algorithm (IWO)	Image datasets
Kumar et al. [454]	LSTM			✓		Artificial bee colony (ABC)	Integrating big data driven sentiments polarity
Das et al. [455]	RNN		✓	✓		Flower pollination (FP) algorithm	Modeling of electron Beam welding process
Gong et al. [456]	LSTM		✓	✓	✓	Fireworks Algorithm (FWA)	Air-conditioning load data of a union office
Chen et al. [457]	LSTM	✓	✓			Hybrid coding particle swarm optimization (HCPSO)	Series prediction and Nonlinear system identification

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Bacanin et al. [458]	CNN			✓		Firefly algorithm (FA)	Medical image classification (IXI and cancer dataset)
Sherly and Jaya [459]	CNN			✓		Improved firefly algorithm (IFA)	Scene character recognition
Datta And Chakrabarti [460]	RNN		✓			Fire fly-oriented multi-verse optimizer (FF-MVO)	Classification problem
Alenazy and Alqahtani [461]	DBN	✓				Gravitational search algorithm (GSA)	Facial expression recognition (FER)
Sudha and alarmathi [462]	DBN		✓			Interactive autodidactic school (IAS)	Classification problem
Jammalamadaka and Parveen [463]	DBN	✓				Search and rescue (SAR) algorithm	Classification problem
Gadekallu et al. [464]	CNN			✓		Crow search algorithm (CSA)	Classification: Human–computer interaction (HCI)
Irmak [465]	CNN			✓		Grid search optimization (GSO)	Medical image classification
Arjunagi and Patil [466]	CNN		✓			Adaptive spider monkey optimization (AOSMO)	Identifying and diagnosing maize leaf diseases
Li et al. [467]	RNN		✓	✓		Adaptive dynamic particle swarm optimization (ADPSO)	Air Quality Index Prediction
Oyelade and Ezugwu [468]	CNN	✓		✓		Multiverse optimizer (MVO), SBO & LCBO	Medical image classification

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Tripathi and Maktedar [469]	CNN		✓			Lion assisted firefly algorithm (LA-FF)	Classification problem
Karuppusamy et al. [470]	DBN	✓				Chronological salp swarm algorithm (CSSA)	Intrusion detection system Intrusion detection in cloud
Priya and Chacko [471]	CNN	✓		✓		Improved particle swarm optimized (IPSO)	Medical image classification
Danesh and Vasuhi [472]	CNN	✓				Glow worm swarm optimization (GWSO)	Spectrum sensing ranks
Zhang et al. [473]	LSTM		✓	✓		Genetic algorithm (GA)	Upper Limb Activities Recognition
Farrag et al. [474]	LSTM		✓	✓		Genetic algorithm (GA)	South Australia State (SA) power system
Arora et al. [475]	DAR			✓		Grasshopper optimisation algorithm (GOA)	Wind Power Forecasting
Goay et al. [476]	CNN-LSTM			✓		Adaptive successive halving Optimization (ASH-HPO)	Transient simulations of high-speed channels
Liu et al. [477]	LSTM		✓		✓	Adaptive particle swarm optimization (AHMPSO)	Monitoring of wastewater treatment plant (WWTP)
Davoudi and Thulasiraman [478]	CNN	✓				Genetic algorithm (GA)	Breast cancer classification problem

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Li et al. [478]	DBN	✓	✓			Simulated annealing cuckoo search algorithm (SA-CSA)	Fault diagnosis of railway freight car wheelset
Liu et al. [479]	CNN	✓				Continuous particle swarm optimization (CPSO)	Hyperspectral Image Classification
Brodzicki et al. [480]	DNN			✓		Whale optimization algorithm (WOA)	Classification Dataset (MNIST)
Baniasadi et al. [481]	CNN	✓				Improved particle swarm optimization (NSBPSO)	Intrusion Detection in IoT Systems
Paul et al. [482]	LSTM-DBN			✓		Sparrow search optimization (SSO)	Water quality index prediction
Gonçalves et al. [483]	CNN		✓	✓	✓	Genetic algorithm (GA) & PSO	Cancer detection
Glaret subin and Muthukannan [484]	CNN		✓	✓	✓	Flower pollination optimization algorithm (FPOA)	Multiple eye disease detection
Xu et al. [485]	LSTM			✓		Particle swarm optimization (PSO)	Hydrology (Flood forecasting)
Antony Raj and Giftson Samuel [486]	DRBFNN	✓				Boosted salp swarm optimization (BSSO)	PhotoVoltaic (PV) systems
Hassanzadeh et al. [487]	CNN		✓	✓	✓	Genetic algorithm (GA)	Classification (CIFAR10, MNIST, and EMNIST)

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Palaniswamy [488]	CNN			✓		Swallow swarm optimization (SSO)	Automated bone age assessment and classification
Jalali et al. [489]	CNN			✓		Grey wolf optimization (GWO) algorithm	Wind power forecasting
Lokku et al. [490]	CNN		✓	✓		Fitness sorted rider optimization (FS-ROA)	Face recognition
Ewees et al. [491]	LSTM			✓		Heap-based optimizer (HBO) algorithm	wind power forecasting
Huo et al. [492]	TCN-LSTM	✓				Particle swarm optimization (PSO)	Prediction of reservoir key parameters
Li et al. [493]	CNN-LSTM		✓	✓	✓	Particle swarm optimization (PSO)	Reservoir production prediction
Ge et al. [494]	DBN		✓	✓	✓	Whale optimization algorithm (WOA)	Safety prediction of shield tunnel construction
Kanna and Santhi [388]	CNN-LSTM	✓				Black widow optimization (BWO)	Intrusion Detection Systems
Jalali et al. [495]	CNN		✓	✓	✓	Modified competitive swarm Optimizer (MCSO)	X-ray image based COVID-19 detection

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Li et al. [496]	LSTM		✓	✓	✓	Grey wolf optimization (GWO)	Wind speed forecasting
Michael Mahesh et al. [497]	CNN			✓		Rider border collie optimization (RBCO)	Road intersection classification
Mohakud and Dash [498]	CNN			✓		Grey wolf optimization (GWO)	Medical image classification
Ahmad et al. [499]	DRaNN			✓		Particle swarm optimization (PSO)	Intrusion detection in the industrial internet of things
Chen et al. [500]	CNN			✓		Chimp optimization algorithm (ChOA)	Diagnose Parkinson's disease
Karthiga et al. [501]	CNN		✓	✓	✓	Grey wolf optimization (GWO) & ABC	Biomedical Signal Processing and Control
Kanipriya et al. [502]	CNN-LSTM		✓	✓	✓	Improved capuchin search algorithm (ICapSA)	Malignant lung nodule detection
Hu et al. [503]	LSTM		✓		✓	Grasshopper optimization algorithm (GOA)	Building Engineering
Raziania and Azimbagirad [504]	CNN			✓		Moth flame optimization (MFO)	Sensor-based human activity recognition

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Falahzadeh et al. [505]	CNN			✓		Grey wolf optimization (GWO)	Speech Emotion Recognition
Vigneshwaran et al. [506]	CNN			✓	✓	Particle swarm optimization (PSO)	Recognition of partial discharge (PD)
Jalali et al. [507]	LSTM			✓	✓	Grasshopper optimization algorithm (GOA)	wind speed forecasting
Surya and Senthilselvi [508]	LSTM			✓		Seagull optimization algorithm (SOA)	Identification of oil authenticity and adulteration
Balasubramanian et al. [509]	CNN		✓	✓		Particle swarm optimization (PSO)	Medical image classification
Pandey and Kamal Jain [510]	CNN			✓	✓	Opposition-based symbiotic organisms search (OSOS)	Medical image classification
Challapalli and Devarakonda [511]	CNN		✓	✓		Hybrid particle swarm grey wolf (HPSGW)	Classification of Indian classical dances
Rodrigues et al. [512]	CNN			✓	✓	Genetic algorithm (GA)	Medical image classification—MRI images
Sasank and Venkateswarlu [513]	CNN	✓				Adaptive rain optimizer algorithm (AROA)	Medical image classification
Kavitha and Prasad [514]	CNN	✓				Sand piper optimization (SPO) Algorithm	Medical image classification
Qader et al. [515]	CNN	✓	✓	✓	✓	Improved harris hawks optimization (HHO)	Medical image classification (brain tumor)
Karthik and Sethukarasi [516]	DBM			✓		Hybrid atom search arithmetic optimization (HASAO)	Natural language processing

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Li et al. [517]	LSTM		✓	✓	✓	Grey wolf optimization (GWO)	Water resources management
Gaurav et al. [518]	CNN	✓				Hosted cuckoo optimization (HCO)	Speaker identification framework
Kaushik et al. [519]	DBN	✓				Whale optimization algorithm (WOA)	Software development effort estimation
Liu et al. [520]	LSTM		✓	✓		Particle swarm optimization (PSO)	Short-term subway inbound passenger flow prediction
Souissi and Ghorbel [521]	LSTM		✓			Genetic algorithm (GA)	Click-through rate prediction-digital advertising industry
Balasubramanian et al. [522]	DBN			✓		Salp swarm optimization algorithm (SSA)	Medical image classification
Mukherjee et al. [523]	CNN			✓	✓	Grey wolf optimization (GWO)	Identification of the types of disease
Ponmalar and Dhanakoti [524]	CNN		✓	✓		Hybrid whale tabu optimization (HWTO)	Intrusion detection in big data
Suresh et al. [525]	RNN	✓				Flamingo search optimization (FSO)	Disease diagnosis
Xu et al. [526]	LSTM		✓	✓	✓	Whale optimization algorithm (WOA)	Short-term traffic flow prediction

Table 3 (continued)

Authors & dates	The deep learning categories	Optimized components: 1. Weights, 2. Layers & Nodes 3. Other Hyper parameters 4. Learning parameters & Activation function				The meta-heuristic algorithm used for training deep learning	Application / dataset
		1	2	3	4		
Tuerxun et al. [527]	LSTM		✓	✓	✓	Modified tuna swarm optimization (MTSO)	Wind speed prediction
Chandraraju and Jeyaprakash [528]	DBN	✓				Chaotic Krill Herd optimization (CKHO)	Diagnosis of breast abnormalities
Jiang et al. [529]	CNN-LSTM			✓		Improved whale optimization algorithm (IWOA)	A Fault Feature Extraction
Fetanat et al. [530]	CNN-FENN			✓	✓	Improved Harris Hawks optimization (IHHO)	Medical image classification
Jiang et al. [531]	LSTM		✓		✓	Sine–Cosine algorithm (SCA-HHO)	Ship attitude prediction
Gampala et al. [532]	DBN	✓				Hosted cuckoo optimization algorithm (HO-COA)	Diagnosis of COVID-19
Li et al. [533]	DBN		✓			Particle swarm optimization (PSO)	Product quality monitoring
Yu et al. [534]	CNN			✓	✓	Enhanced chicken swarm algorithm (ECSA)	Crack detection of concrete structures
Li et al. [535]	CNN			✓	✓	Multi-strategy particle swarm optimization (MSPSO)	Fault diagnosis method for aircraft EHA
Pellegrino et al. [536]	DNN		✓		✓	Particle swarm optimization (PSO) & GA	Predicting BRCA1/BRCA2 Pathogenicity
Mohapatra et al. [537]	CNN	✓				Cat swarm updated black widow (CSUBW)	Medical image classification
Ragab et al. [389]	CNN			✓		Enhanced gravitational search optimization (EGSO)	COVID-19 diagnosis
Shankar et al. [538]	RNN	✓	✓	✓	✓	Aquila optimization algorithm (AOA)	Fruit classification
Fan et al. [539]	CNN		✓	✓	✓	Hybrid Sparrow Search Algorithm (HSSA)	Image classification

4.2.1 Investigation of optimized components in DL architectures

As an optimization problem, MH algorithms formulate the optimal estimation of DL components (such as hyper-parameter, weights, number of layers, number of neurons, learning rate, etc.). This section examines the abundance of MH use for optimized components in DL architectures (according to the papers in Table 3). Figure 20 represents the relative abundance of research on optimized components in DLs using MH algorithms. As demonstrated in Fig. 20, in 61 studies (20%), weights and biases have been adjusted using MH algorithms. In 76 studies (26%), the number of layers and neurons in the layers have been adjusted using MH algorithms. Moreover, in 114 studies (38%), hyper-parameters in DL architectures have been adjusted. Finally, in 47 studies (16%), learning parameters, learning algorithms or activation functions have been adjusted.

Figure 21 also indicates the relative abundance of research in the simultaneous optimization of two components of DLs. As can be seen in Fig. 21, in 14 studies, weights and layers, and neurons were adjusted simultaneously. In 12 studies, weights and hyper-parameter; in 4 studies, weights and learning parameters; in 40 studies, the number of layers and number of

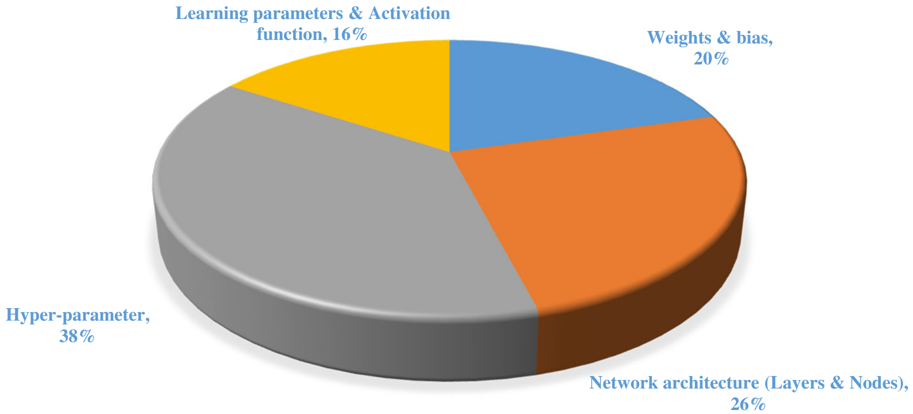

Fig. 20 Relative abundance of research on optimized components in DL architectures using MH algorithms

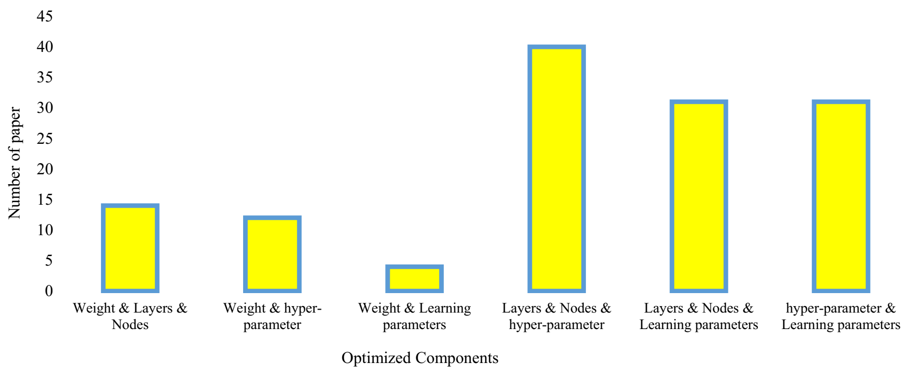

Fig. 21 Relative abundance of research in the simultaneous optimization of two components of DL using MHs

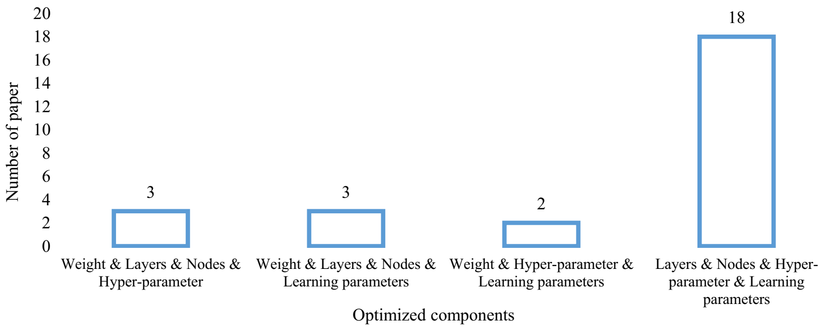

Fig. 22 Relative abundance of research in the simultaneous optimization of three components of DL using MHs

neurons and hyper-parameter; in 31 studies, the number of layers and number of neurons and learning parameters, and in 31 studies hyper-parameter and learning parameters have been adjusted simultaneously. Figure 22 also represents the relative abundance of research in the simultaneous optimization of three DL components (according to Table 3).

As can be seen, in 3 studies, weights, the number of layers and number of neurons and the hyper-parameter were adjusted simultaneously. In 3 studies, weights, number of layers and number of neurons and learning parameters; in 2 studies, weights, hyper-parameter and learning parameters; in 18 studies, hyper-parameter, number of layers and number of neurons and learning parameters were adjusted simultaneously. According to Table 3, in only 2 studies, all four DL components were adjusted simultaneously. Therefore, very little research has been done in this area (simultaneous optimization of three/four components).

4.2.2 Investigation of Meta-Heuristic Algorithms Used in DL's Optimization

According to Table 3, many MH algorithms have been developed to optimize DL architectures. Figure 23 represents the MH algorithms used to optimize DLs. PSO with 48 implementations and GA with 27 implementations were the most used algorithms. EA,

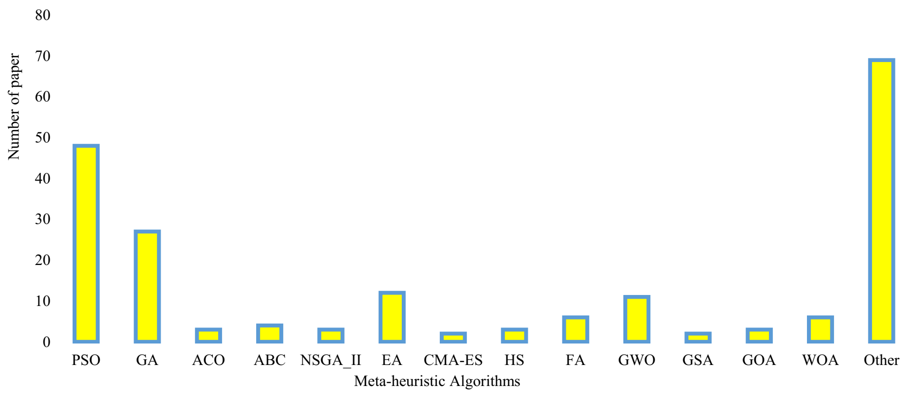

Fig. 23 Meta-heuristic algorithms used in DL's optimization

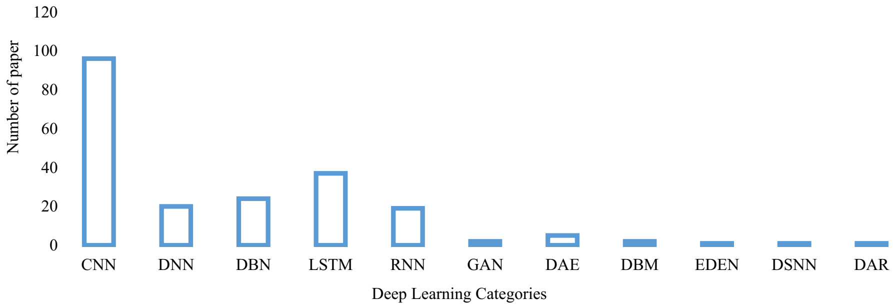

Fig. 24 The abundance of MHs used for different types of DL architectures

GWO, FA, WOA, ABC, ACO, HS, NSGA_II, CMA-ES, and GOA algorithms are also in the next ranks.

4.2.3 Investigating the Abundance of MHs Used for Different Types of DL Architectures

Some of the popular DL architectures are Long short-term memory (LSTM), Convolutional Neural Networks (CNNs), Deep Belief Networks (DBN), Recurrent Neural Networks (RNN), Deep Boltzmann Machines (DBM), Deep Auto Encoder (DAE), and Deep Neural Networks (DNN). In this section, the abundance of MHs used for different DL architectures is investigated (Fig. 24). CNN with 96 implementations, LSTM with 37 implementations, and DBN with 24 implementations were the most used DL architectures, which are set using MH algorithms. DNN, RNN, DAE, DBM, GAN, DSNN, DAR, and EDEN architectures are also in the next ranks.

4.2.4 Checking the Number of Papers Published in Journals and Years

In this section, the papers in Table 3 are categorized according to the type of journals and the year of their publication. Figure 25 demonstrates the percentage of papers published in various journals (based on Table 3). As indicated, 71 papers (37%) in Elsevier, 39 papers (20%) in

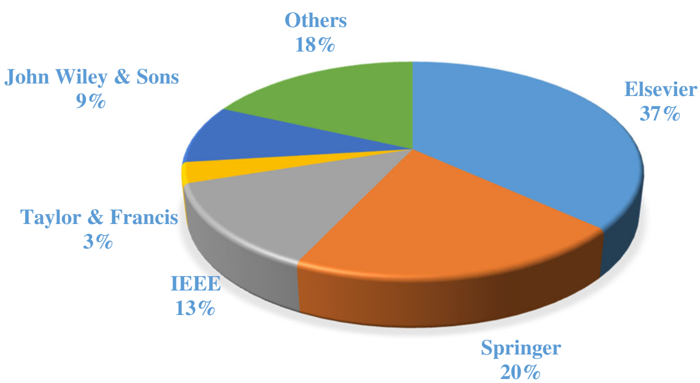

Fig. 25 Papers published in journals (based on Table 3)

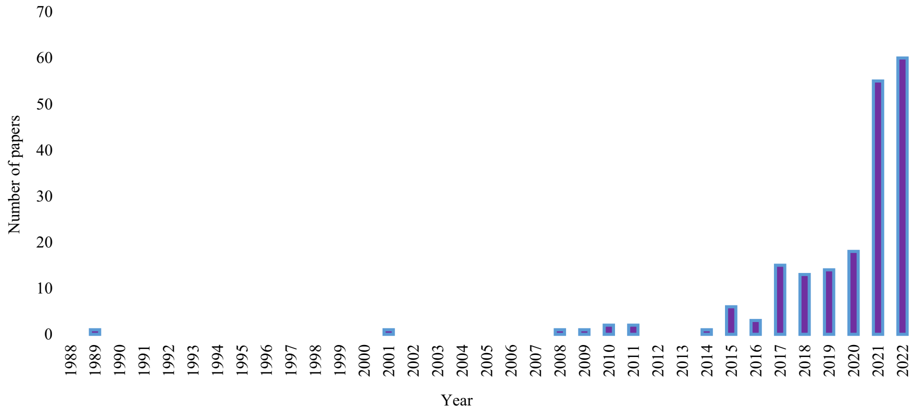

Fig. 26 Changes in the number of papers published in different years about the use of MH for Training DLs

Springer, 25 papers (13%) in IEEE, 6 papers (3%) in Taylor & Francis, and 17 papers (9%) In John Wiley & Sons, and 35 papers (18%) in other journals have been published regarding the use of MH for DL architectures.

Figure 26 also represents the changes in the number of papers published in different years about the use of MH for Training DLs. Between 1988 and 2016, few papers were developed for DL optimization. From 2017 to 2020, DL optimization received a little more attention from researchers, and the number of papers in this field increased. But from 2021 to 2022, many researchers have worked on DL optimization. This problem is still a challenge, and many problems need to be resolved.

4.2.5 Applications of DLs

In this section, the application of the papers in Table 3 is evaluated. Figure 27 shows the application of the papers regarding the use of MH for DLs. 48 papers in medical image classification and medical diseases diagnosis, 46 papers in Benchmark problem (Classification, prediction, time series, optimization, recognition, system identification), 44 papers in electrical engineering, signal processing and energy systems, 23 papers in civil engineering and environmental management, 8 papers in mechanical engineering, 3 papers in biomedical and chemical engineering, 4 papers in economy and product quality, and 17 papers in other applications have been published regarding the use of MH for ANNs.

As can be seen, most of the DLs were implemented on medical image classification and benchmark problems (such as MNIST, CIFAR-10, Caltech, CINIC-10, and EMNIST datasets). According to Table 3, evolutionary CNN architectures have been used in many medical image classification applications. The results of these papers show that the proposed hybrid DL architectures perform better than others. Therefore, the combination of MH and CNNs methods can be useful for medical applications.

4.2.6 Contributions of Different Continents in Using the Hybrid MH-DL Models

Figure 28 shows the distribution of studied papers according to the affiliation of the authors for each continent. As can be seen, Asia has the largest portion of contributions in the world, while America has the lowest contributions.

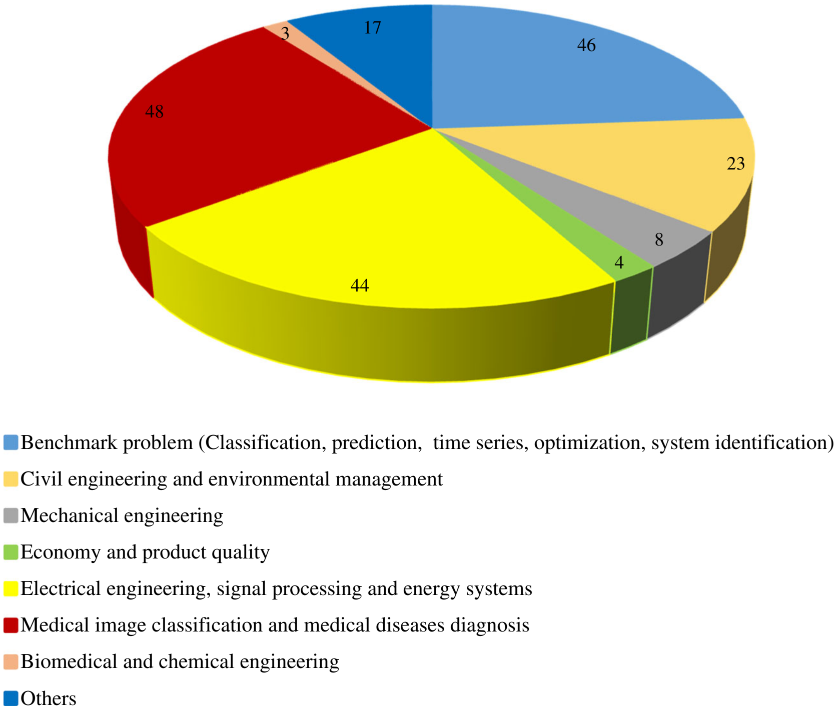

Fig. 27 Application of papers regarding the use of MH for DLs

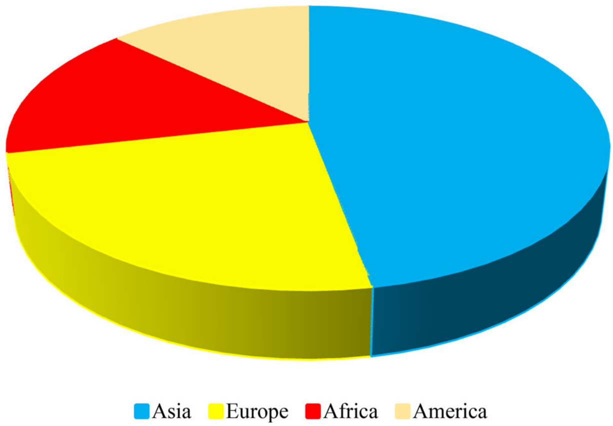

Fig. 28 Contributions of different continents in using the hybrid MH-DL models

5 Discussion, Statistical Results, Limitations, and Future Challenges

5.1 Discussion and Statistical Results of Tables 2 and 3

As can be seen from the results of Tables 2 and 3, neural network optimization has been considered by researchers from the past to the present. But the optimization of DL parameters has recently been considered, and more research is needed in this field. The main reason is that the DL concept has been seriously pursued since 2008. Therefore, many challenges and more research are needed in this field. The existence of many parameters in DL architectures has led to the use of MH algorithms to optimize them. According to Table 3, DL optimization has been considered by researchers since 2015.

According to the literature review, well-known MH algorithms such as GA and PSO have been used for training the NN and DL. But according to the No Free Lunch (NFL) theorem, each problem has its characteristics, and different algorithms must be tested to solve it [540]. According to the NFL theorem, it is very difficult to find a comprehensive MH algorithm to solve various problems [541]. Therefore, an MH algorithm may not be suitable for optimizing the NN and DL parameters. However, it works well in solving some problems. In addition, the only way to determine the convergence of the MH algorithm is through its experimental evaluations. Because MH algorithms search the problem space (based on their operators), it is difficult to choose the MH algorithm as the best method for a particular problem. Therefore, it is necessary to use different algorithms to optimize the NN and DL parameters.

In many research studies on optimization problems [18, 19, 542, 543], improving common versions of MH algorithms (and combination of algorithm) has increased exploitation and exploration power. In some recent research [66, 67, 120], new MH algorithms have been introduced, which have performed better than the old algorithms in many optimization problems. According to the literature review (Tables 2 and 3), in most research, common algorithms (such as PSO and GA) have been used to optimize NN and DL. Therefore, the development of old MH algorithms, as well as novel MH algorithms for optimizing NN and DL parameters, is a new challenge, which can be seen in recent papers in Tables 2 and 3.

It is complicated to find the best possible solution in the search space in large-scale optimization problems. Moreover, changing algorithm variables does not have much influence on the algorithm convergence. Therefore, for massive dataset with high complexity, even if the researchers have determined accurate initial parameters, the algorithm will not be able to perform adequate exploration and exploitation. Consequently, to achieve comprehensive global and local searches, we need to apply powerful operators to make better exploration and exploitation. MH algorithms can be combined with others and overcome this problem by using the advantages and operators of other algorithms. In recent decades, researchers have utilized a combination of algorithms to improve the performance of the optimization process. The weakness of an algorithm can be compensated by the operation of other algorithms.

Most researchers tend to extend novel hybrid algorithms by combining MHs to optimize the hyper-parameters of DLs and ANNs. The development of hybrid MHs helps improving algorithms performance and capable of solving complex optimization problems. According to the results, many researches have used the modification and hybridization of meta-heuristic algorithms to optimize ANN and DL parameters. Also, the performance of the proposed hybrid MH algorithms have been better than others.

In general, the optimal performance of the MHs should be able to achieve a suitable trade-off between exploration and exploitation features. The exploration operator can explore the search space more efficiently and perform a global search to avoid getting stuck in local minimum, but it may encounter slow convergence. On other hand, the exploitation operator leads to very high convergence rates, but may be trapped in a local minimum. Among the existing MH algorithms, some of them are better in convergence trend (exploitation) while others have more ability to avoid getting trapped in local optimum (exploration). Table 4 indicates the comparison of different MH algorithms in terms of their ability of finding global optimum, convergence trend, exploitation ability, exploration ability, parameter setting, and implementation. As can be seen, grey wolf optimizer, black widow optimization, chimp optimization algorithm, differential evolution, red fox optimization, capuchin search algorithm, and gannet optimization algorithm perform well in most properties and their operators can be used to improve other architectures. This framework is useful for researchers for their applications in improved hybrid algorithm.

According to the statistical results of Table 2, in only one study, the simultaneous optimization of all components (weights, number of layers, number of neurons and learning functions/parameters) of neural networks has been investigated. Also, in two study, the simultaneous optimization of all components (weights, number of layers and neurons, hyper-parameter, and learning functions/parameters) of DLs has been investigated. However, there is no research on training DL (simultaneous optimization of all components). So researchers in the future can optimize all components simultaneously to improve network performance. This is a challenge for both neural networks and DL architectures. In addition, in neural networks, in most cases, the weight of the network is optimized. But in DL architectures, weight, hyper-parameter, and network structure are optimized equally. Since optimizing ANN and DL architectures is a complex and multi-objective problem (MOO), using multi-objective MH algorithms or developing new multi-objective MH algorithms is also challenging. While in very few papers, multi-objective MH algorithms have been used to optimize ANN and DL parameters (as represented in Tables 2 and 3).

In optimizing DL algorithms, CNN architecture is more trained. According to the NFL theorem for MH algorithms, implementing all DL algorithms for various problems is also challenging. In fact, different DL architectures need to be implemented for different problems and their experimental results evaluated. Therefore, optimizing other DL architectures can be considered to solve various problems in the future. Table 5 also indicates the advantages and disadvantages of compared techniques.

5.2 Limitations of Deep Learning

Notwithstanding the positive outcomes of the reviewed papers, there are still some challenges and limitations related to deep learning and DL methods that should be addressed.

- *Over-fitting problem in a deep neural network* Many parameters relate to unseen datasets in some complex applications. This can cause a difference in the error caused by the training dataset and the new unseen dataset.
- *Hyper-parameters optimization* DL architectures have several hyper-parameters, for example, learning rate, number of hidden layers, number of neurons in each hidden layer, number of convolution and max-pooling layers, and so on. Most often these hyper-parameters are adjusted by trial and error method. MH algorithms formulate the optimal estimation of DL components (such as hyper-parameter, weights, number of layers, number of neurons, learning rate, etc.).
- *Computing Power Required* High computing power is required to tackle a real-world problem using DL models. Therefore, experts are trying to develop high-performance multi-core GPUs and similar processing units such as TPUs in the future.

Table 4 Comparison of MH algorithms in different criteria

MH algorithm	Exploitation ability	Exploration ability	Convergence trend	Ability of finding global optimum	Parameter setting	Implementation
Genetic algorithm	Medium	High	Very slow	Low	Medium	Simple
Particle swarm optimization	Very high	Low	Fast	Low	Medium	Simple
Simulated annealing	Medium	Medium	Very slow	Low	Easy	Simple
Differential evolution	High	Medium	Very fast	High	Easy	Medium
Artificial bee colony	High	Medium	Medium	Medium	Easy	Simple
Ant colony optimization	High	Medium	Fast	High	Hard	Medium
Tabu search (TS)	Low	High	Slow	Medium	Medium	Medium
Biogeography-based optimization	High	High	Very fast	High	Medium	Medium
Whale optimization algorithm	High	Medium	Medium	High	Easy	Medium
Gravitational search algorithm	High	Medium	Medium	Very high	Medium	Simple
Grasshopper optimization algorithm	Medium	High	Medium	High	Medium	Simple
Cuckoo search	High	Medium	Medium	Very high	Easy	Medium
Firefly algorithm	Medium	Low	Medium	Very low	Easy	Easy
Grey wolf optimizer	Very high	High	Very fast	Very high	Easy	Medium
Harmony search	High	High	Very fast	High	Easy	Simple
Interior search algorithm	Medium	High	Fast	Medium	Medium	Simple
Salp swarm algorithm	Medium	High	Medium	High	Easy	Medium

Table 4 (continued)

MH algorithm	Exploitation ability	Exploration ability	Convergence trend	Ability of finding global optimum	Parameter setting	Implementation
Weighted superposition attraction	Very high	Medium	Fast	High	Medium	Complex
Black widow optimization	Very high	High	Very fast	High	Easy	Medium
Chimp optimization algorithm	Very high	Very high	Very fast	High	Easy	Medium
Red fox optimization	Very high	Very high	Fast	Very high	Easy	Medium

Table 5 Advantages and disadvantages of compared DL techniques

DL method	Advantages	Disadvantages
DNN	1. Its implementation is simple. Deep neural networks with multiple hidden layers automatically discover the features of complex objects such as images 2. ANNs can be applied in parallel and work fast. Consequently, they are specially programmed to perform online processes 3. It is unnecessary to identify key criteria where DNN can define all criteria and then determine which criteria are relevant 4. DANN implementations allow developers to add learning capabilities to their applications 5. Self-organization and Usability in big data due to the training process	1. Lack of sufficient theoretical foundation 2. Computationally cost. It requires a long training time. Learning a DNN when dealing with big data can take days or months 3. In DNNs, a large number of hyper-parameters need to be adjusted. Moreover, with an increasing number of hidden layers and nodes, the training algorithm is more likely get trapped in the local optimal 4. A large amount of training data is required to training process
DBN	1. The training of DBNs is divided into two phases: the pre-training and the fine-tuning. In the pre-training process, an unsupervised algorithm based training is performed for the feature extraction; while in the fine-tuning process, a supervised algorithm is performed for further adjustment of the hyper-parameters 2. DBN networks have a level of flexibility 3. DBN is applied to applications with unlabeled data. Moreover, the overfitting and underfitting errors can be avoided	1. Deep in time (two phases learning) 2. local information (Spatial data) is lost as the network gets deeper
CNN	1. CNN is the first truly successful DL method due to the successful training of the hierarchical layers 2. CNN requires minimal pre-processing 3. It is suitable for feature extraction, image classification, image recognition, and prediction problems 4. CNN reduce the number of parameters by leverages spatial relationships 5. CNN Fine-tunes all the layers of the network	1. A large amount of training data is required to training process 2. It requires a lot of time and computing resources
RNN	1. RNNs Deal with sequential data 2. RNNs can capture longer context patterns 3. RNNs are used to earn metal	1. It requires a long training time 2. Training process is difficult 3. The performance of RNN decreases rapidly

Table 5 (continued)

DL method	Advantages	Disadvantages
LSTM	1. It allows information to flow in both forwards and backward processes within the network 2. It has a sensible processing for time series data 3. It can learn its tasks without ability to predict the local sequence	1. Training process is difficult 2. Complex network structure 3. It is computationally expensive
DBM	1. Able to learn internal representations 2. It is a fully connected NN 3. DBM Deals strongly with ambiguous inputs	1. It requires a long training time 2. Difficult to train
DAE	1. It has the ability to extract useful features during the propagation and filter the useless data 2. DAE is an unsupervised DL architecture used for dimensionality reduction	1. Training process is difficult 2. DAE Requires pre-training

- *Gradient-based learning* The learning process of DL architectures is considered one of the most challenging machine learning problems. Several past studies have used gradient-based methods to train DL architectures. However, gradient-based methods have major drawbacks such as stucking at local minimums in multi-objective cost functions, expensive execution time due to calculating gradient information with thousands of iterations and needing the cost functions to be continuous. Since training the ANNs and DLs is an NP-hard optimization problem, their structure and parameters optimization using the meta-heuristic algorithms has been considerably raised.
- *Dataset unavailability for various applications* DL requires a large amount of training dataset. The classification accuracy of the DL architectures is highly dependent on the quality and size of the dataset. However, unavailability of the dataset is one the biggest barrier in the success of DL architectures.
- *Determining the type of DL architecture to solve a particular problem* Many studies have used different DL architectures to solve engineering and medical problems. However, there is no explanation for how these architectures are chosen to solve specific problems.
- *Heterogeneity in image dataset* The nature of data varies from hardware to hardware and thus, there are many variations in images due to sensors and other factors. In addition, the wide range of medical applications requires the combination of several different datasets for learning and accuracy of algorithms.
- *Architecture Implementation Cost* Feature extraction can be done in advance and then the proper methods can be implemented. The purpose of this process is to reduce the computing runtime (training) and computing power required.
- *Lack of results of different DL architectures on benchmark database* The lack of results of different DL architectures is still a challenge in solving many benchmark database or benchmark engineering problems. For example, in some studies [544, 545], the authors have used different DL architectures and compared the results with the decision tree.

- *Reasonable Computing Time* Some applications with many variables in some deep learning methods, (such as DNN) have high dimensions, which poses a challenge for these models to obtain an accurate DNN in a reasonable execution time.
- *One-Shot Learning* DL architectures require a lot of training data to provide high-quality results. For example, the Image-Net database contains more than a million images, and the DL architecture often requires thousands of instances to classify them correctly. Human does not need thousands of bicycle images to learn a picture of a bicycle. When a bicycle is shown to a child, they can often recognize another bicycle, even in different models, shapes, and colors.
- *Imbalanced data* In this problem, one or more classes may have very few representatives in the training process. MH algorithms can be used to deal with such problems.
- *Theoretical backbone* Unlike decision trees, SVMs, and other machine learning architectures, most of the DL methods are yet to possess a strong theoretical backbone.

5.3 Future Work

While deep learning models have been successfully applied in various application fields, there are future works and challenges that require to be addressed. Scientists and researchers should do more research and work to overcome the challenges facing the future of deep learning. In addition, more DL techniques and inspirations are needed to develop new DL architectures. New techniques will be necessary for complex applications. In addition, DL architectures can take advantage of various sub-domains of swarm intelligence and evolutionary computation that are still unexplored. In this section, according to the literature review, some relevant perspectives for future work are listed.

- *Design of DL methods* Deep learning is used as an efficient method to deal with big data problem. Furthermore, DL method has get great success with a large number of unlabeled data. However, rather strong techniques are required when a limited training data is available. Therefore, it is important to consider designing DL techniques from multiple training datasets in the future.
- *DL and mobile devices* The idea of DL chips has attracted the attention of many researchers. Deep learning techniques can be implemented in mobile devices with low-power energy.
- *Transfer Learning* The learning architecture in the human brain has evolved over millions of years and has been transferred from generation to generation. Humans transfer part of their learning as an experience to future generations. In addition, humans constantly learn about different tasks that help them learn specific tasks faster. For this reason, learning different problems is achieved by making basic and easy settings. Developing the concept of transfer learning in DL is one of the challenges in this field and can be a new field of work for researchers in the future. Transfer learning reduces training time and the use of previous learning experiences in new tasks.
- *DL and Reinforcement Learning (RL)* RL mainly involves goal-oriented algorithms that learn how to achieve a complex goal. Recently, the combination of DL and RL methods has attracted the attention of researchers. These methods have led to several applications such as self-driving cars and AlphaGo. Future works can focus on exploring MH algorithms in optimizing learning methods in deep RL.
- *Unsupervised Learning-Based DL* Because having labeled data is usually costly, the next generation of DL techniques is more semi-supervised and unsupervised. Here, clustering concepts and algorithms can be used to improve the performance of DL algorithms.

- *Stability of DL* Stability analysis of DL is considered an important problem in this field due to its numerous advantages for different applications. Therefore, we should focus on some problems such as stability analysis, state estimation, and synchronization for DLs.
- *Dimensionality reduction* This problem is one of the most prevalent challenges needed to be addressed since the number of the features from deep learning method can be huge. This problem weakens the performance of the algorithm, since most of these features are redundant. To address this problem in the future, various MHs can be combined with DL models. MH algorithms first select the optimal features and then transfer them to a DL model.
- *Developing more challenging evolutionary DL models* There are many papers in this field (EvoDL), but not much paper has been undertaken to evolve Generative Adversarial Network (GAN) by using MH algorithms. In addition, MH-based optimization algorithms may also be explored to evolve DL extensions of non-iterative learning paradigms.
- *Energy-efficient Learning Problem* In most cases, DL architectures that work on big data are inefficient in energy consumption. On the other hand, the human brain requires very little energy to learning and often does not perform accurate calculations (estimates). This energy is enough to learn about many problems and can add to the power of generalization. Therefore, in the future, DL architectures must be designed to be energy efficient.
- *Improvement of MHs* MH algorithms still need to be improved before applying them to the deep learning architecture. Since most of MHs have a high capability in exploration or exploitation, it is a challenging work to detect the MH that can balance between exploration and exploitation. Furthermore, many of the MH algorithms ranked in CEC competitions have not been used to optimize parameters of DLs.

6 Conclusions

Deep learning is a new approach to machine learning in recent years and has been successfully applied in various applications. DL techniques are superior to traditional ML algorithms due to data availability and systems processing power development. With the advent of the big data era, much faster data collection, storage, updating, and management advances have become possible. In addition, the development of GPU has made efficient processing in large data sets. These dramatic advances have led to recent advances in DL techniques. DL methods have been used in various applications, including image classification, prediction, Phoneme recognition, hand-written digit recognition, etc.

The learning process and hyper-parameter optimization of ANNs and DLs is considered one of the most difficult machines learning challenges and has recently attracted many researchers. Training the ANNs and DLs is an NP-hard optimization problem with several theoretical and computational limitations. MH algorithms formulate NN and DL components as an optimization problem. Therefore, this research presents a comprehensive review of NNs and DLs' optimization using meta-heuristic algorithms.

As can be seen from the results, neural network optimization has been considered by researchers from the past to the present. But the optimization of DL parameters has recently been considered. According to the literature review, well-known MH algorithms have been used for training the NN and DL. Therefore, the development of these algorithms, as well as novel MH algorithms for optimizing NN and DL parameters, is a new challenge. According to the statistical results, researchers can optimize all components of ANNs and DL architectures simultaneously to improve network performance in the future. In this way, they can use multi-objective algorithms to teach architectures better. According to the results, evolutionary CNN architectures have been used in many medical image classification applications. The results of these papers show that the proposed hybrid MH-CNN architectures perform better than others. Therefore, the combination of MH and CNNs can be useful for medical applications. In most papers, MHs have been used for image classification problems. Therefore, there is still room to apply these hybrid methods in different applications and evaluate their performance on different challenging real-world datasets.

In this paper, we have reviewed the latest developments in the use of MH algorithms in the DL methods, presented their disadvantages and advantages, and pointed out some research directions to fill the gaps between MHs and DL methods. Moreover, it has been explained that the evolutionary hybrid architecture still has limited applicability in the literature. Using MH algorithms to train DLs improves the learning process. This increases the accuracy of the algorithm and reduces its execution time. The combination of MH and DLs provides a good start to the DL process and improves the DL performance. It is difficult to assess whether the deep learning methods will be at the academic boundary (without the integration with MH). It is expected that in the coming years, combining DL with MH will accelerate the training process and maintain high performance. According to the review of papers, using MH algorithms to optimize DL architectures is still challenging, and more research is needed in this field. It is expected that MH algorithms will be used more in the coming years to improve the performance of DL architectures. However, relevant publications in this way are still rare.

Acknowledgements Not applicable.

Funding Not applicable.

Availability of data and material The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethical approval This paper does not contain any studies with human participants or animals.

References

1. Skansi S (2018) Introduction to deep Learning: from logical calculus to artificial intelligence. Springer, Cham
2. Aggarwal CC (2018) Neural networks and deep learning. Springer, Cham
3. Bouwmans T, Javed S, Sultana M, Jung SK (2019) Deep neural network concepts for background subtraction: a systematic review and comparative evaluation. Neural Netw 117:8–66
4. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
5. Lanillos P, Oliva D, Philippsen A, Yamashita Y, Nagai Y, Cheng G (2020) A review on neural network models of schizophrenia and autism spectrum disorder. Neural Netw 122:338–363

6. Boveiri HR, Khayami R, Javidan R, MehdiZadeh AR (2020) Medical image registration using deep neural networks: a comprehensive review. arXiv preprint arXiv:2002.03401
7. Lopez-Garcia TB, Coronado-Mendoza A, Domínguez-Navarro JA (2020) Artificial neural networks in microgrids: a review. Eng Appl Artif Intell 95:103894
8. Han F, Jiang J, Ling QH, Su BY (2019) A survey on metaheuristic optimization for random single-hidden layer feedforward neural network. Neurocomputing 335:261–273
9. Ojha VK, Abraham A, Snášel V (2017) Metaheuristic design of feedforward neural networks: a review of two decades of research. Eng Appl Artif Intell 60:97–116
10. Darwish A, Hassanien AE, Das S (2020) A survey of swarm and evolutionary computing approaches for deep learning. Artif Intell Rev 53(3):1767–1812
11. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26
12. Kubat M (2017) An introduction to machine learning. Springer International Publishing AG, Cham
13. Yingwei L, Sundararajan N, Saratchandran P (1997) A sequential learning scheme for function approximation using minimal radial basis function neural networks. Neural Comput 9(2):461–478
14. Ferrari S, Stengel RF (2005) Smooth function approximation using neural networks. IEEE Trans Neural Netw 16(1):24–38
15. Mosavi MR, Kaveh M, Khishe M (2016a) Sonar data set classification using MLP neural network trained by non-linear migration rates BBO. In: The fourth Iranian conference on engineering electromagnetic (ICEEM 2016), pp. 1–5
16. Mosavi MR, Kaveh M, Khishe M, Aghababaee M (2016b) Design and implementation a sonar data set classifier by using MLP NN trained by improved biogeography-based optimization. In: Proceedings of the second national conference on marine technology, pp. 1–6.
17. Mosavi MR, Kaveh M, Khishe M, Aghababaee M (2018) Design and implementation a sonar data set classifier using multi-layer perceptron neural network trained by elephant herding optimization. Iran J Marine Technol 5(1):1–12
18. Kaveh M, Khishe M, Mosavi MR (2019) Design and implementation of a neighborhood search biogeography-based optimization trainer for classifying sonar dataset using multi-layer perceptron neural network. Analog Integr Circuits Signal Process 100(2):405–428
19. Khishe M, Mosavi MR, Kaveh M (2017) Improved migration models of biogeography-based optimization for sonar dataset classification by using neural network. Appl Acoust 118:15–29
20. Zhang GP (2000) Neural networks for classification: a survey. IEEE Trans Syst Man Cybern Part C (Appl Rev) 30(4):451–462
21. Tong DL, Mintram R (2010) Genetic algorithm-neural network (GANN): a study of neural network activation functions and depth of genetic algorithm search applied to feature selection. Int J Mach Learn Cybern 1(1–4):75–87
22. Jiang S, Chin KS, Wang L, Qu G, Tsui KL (2017) Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. Expert Syst Appl 82:216–230
23. Shang L, Huang DS, Du JX, Zheng CH (2006) Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network. Neurocomputing 69(13–15):1782–1786
24. Zhao ZQ, Huang DS, Jia W (2007) Palmprint recognition with 2DPCA+ PCA based on modular neural networks. Neurocomputing 71(1–3):448–454
25. Wang XF, Huang DS, Du JX, Xu H, Heutte L (2008) Classification of plant leaf images with complicated background. Appl Math Comput 205(2):916–926
26. Luo H, Yang Y, Tong B, Wu F, Fan B (2017) Traffic sign recognition using a multi-task convolutional neural network. IEEE Trans Intell Transp Syst 19(4):1100–1111
27. Kaveh M, Mesgari MS, Khosravi A (2020) Solving the local positioning problem using a four-layer artificial neural network. Eng J Geospat Inf Technol 7(4):21–40
28. Hwang JN, Kung SY, Niranjan M, Principe JC (1997) The past, present, and future of neural networks for signal processing. IEEE Signal Process Mag 14(6):28–48
29. Subudhi B, Jena D (2011) Nonlinear system identification using memetic differential evolution trained neural networks. Neurocomputing 74(10):1696–1709
30. Razmjooy N, Ramezani M (2016) Training wavelet neural networks using hybrid particle swarm optimization and gravitational search algorithm for system identification. Int J Mechatron Electr Comput Technol 6(21):2987–2997
31. Gorin A, Mammone RJ (1994) Introduction to the special issue on neural networks for speech processing. IEEE Trans Speech Audio Process 2(1):113–114

32. Khalifa MH, Ammar M, Ouarda W, Alimi AM (2017) Particle swarm optimization for deep learning of convolution neural network. In: 2017 Sudan conference on computer science and information technology (SCCSIT), pp. 1–5
33. Lopez-Rincon A, Tonda A, Elati M, Schwander O, Piwowarski B, Gallinari P (2018) Evolutionary optimization of convolutional neural networks for cancer miRNA biomarkers classification. Appl Soft Comput 65:91–100
34. Dufourq E, Bassett BA (2017) Eden: evolutionary deep networks for efficient machine learning. In: 2017 pattern recognition association of South Africa and robotics and mechatronics (PRASA-RobMech), pp. 110–115
35. Wang B, Sun Y, Xue B, Zhang M (2018) A hybrid differential evolution approach to designing deep convolutional neural networks for image classification. In: Australasian joint conference on artificial intelligence. Springer, Cham, pp 237–250
36. Wang C, Xu C, Yao X, Tao D (2019) Evolutionary generative adversarial networks. IEEE Trans Evol Comput 23(6):921–934
37. Ye F (2017) Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data. PLoS ONE 12(12):e0188746
38. Peng L, Liu S, Liu R, Wang L (2018) Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 162:1301–1314
39. Wang Y, Zhang H, Zhang G (2019) cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks. Swarm Evol Comput 49:114–123
40. Shinozaki T, Watanabe S (2015) Structure discovery of deep neural network based on evolutionary algorithms. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp. 4979–498
41. David OE, Greental I (2014). Genetic algorithms for evolving deep neural networks. In: Proceedings of the companion publication of the 2014 annual conference on genetic and evolutionary computation, pp. 1451–1452
42. Lander S, Shang Y (2015) EvoAE--a new evolutionary method for training autoencoders for deep learning networks. In: 2015 IEEE 39th annual computer software and applications conference, vol. 2, pp. 790–795
43. Rosa G, Papa J, Marana A, Scheirer W, Cox D (2015) Fine-tuning convolutional neural networks using harmony search. In: Iberoamerican congress on pattern recognition, pp. 683–690
44. Rosa G, Papa J, Costa K, Passos L, Pereira C, Yang XS (2016) Learning parameters in deep belief networks through firefly algorithm. In: IAPR workshop on artificial neural networks in pattern recognition, pp. 138–149
45. Martín A, Lara-Cabrera R, Fuentes-Hurtado F, Naranjo V, Camacho D (2018) EvoDeep: a new evolutionary approach for automatic deep neural networks parametrisation. J Parallel Distrib Comput 117:180–191
46. Banharnsakun A (2019) Towards improving the convolutional neural networks for deep learning using the distributed artificial bee colony method. Int J Mach Learn Cybern 10(6):1301–1311
47. Van Der Smagt PP (1994) Minimisation methods for training feedforward neural networks. Neural Netw 7(1):1–11
48. Battiti R (1992) First-and second-order methods for learning: between steepest descent and Newton's method. Neural Comput 4(2):141–166
49. Johnson R, Zhang T (2013) Accelerating stochastic gradient descent using predictive variance reduction. Adv Neural Inf Process Syst 26:315–323
50. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
51. Lan K, Liu L, Li T, Chen Y, Fong S, Marques JAL, Tang R (2020) Multi-view convolutional neural network with leader and long-tail particle swarm optimizer for enhancing heart disease and breast cancer detection. Neural Comput Appl. https://doi.org/10.1007/s00521-020-04769-y
52. Kilicarslan S, Celik M, Sahin Ş (2021) Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. Biomed Signal Process Control 63:102231
53. Son NN, Chinh TM, Anh HPH (2020) Uncertain nonlinear system identification using Jaya-based adaptive neural network. Soft Comput. https://doi.org/10.1007/s00500-020-05006-3
54. Ertuğrul ÖF (2020) A novel clustering method built on random weight artificial neural networks and differential evolution. Soft Comput. https://doi.org/10.1007/s00500-019-04647-3
55. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
56. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
57. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554

58. Basak H, Kundu R, Singh PK, Ijaz MF, Woźniak M, Sarkar R (2022) A union of deep learning and swarm-based optimization for 3D human action recognition. Sci Rep 12(1):1–17
59. Glover F (1986) Future paths for integer programming and links to artificial intelligence. Comput Oper Res 13(5):533–549
60. Holland John H (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
61. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern Part B 26(1):29–41
62. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: MHS'95. Proceedings of the sixth international symposium on micro machine and human science, pp. 39–43
63. Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680
64. Storn R, Price K (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 11(4):341–359
65. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
66. Hayyolalam V, Kazem AAP (2020) Black widow optimization algorithm: A novel meta-heuristic approach for solving engineering optimization problems. Eng Appl Artif Intell 87:103249
67. Khishe M, Mosavi MR (2020) Chimp optimization algorithm. Expert Syst Appl 149:113338
68. Połap D, Woźniak M (2021) Red fox optimization algorithm. Expert Syst Appl 166:114107
69. Pan JS, Zhang LG, Wang RB, Snášel V, Chu SC (2022) Gannet optimization algorithm: A new meta-heuristic algorithm for solving engineering optimization problems. Math Comput Simul 202:343–373
70. Srinivas N, Deb K (1994) Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evol Comput 2(3):221–248
71. Rubinstein RY (1997) Optimization of computer simulation models with rare events. Eur J Oper Res 99(1):89–112
72. Mladenović N, Hansen P (1997) Variable neighborhood search. Comput Oper Res 24(11):1097–1100
73. Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. Evol Comput 9(2):159–195
74. Geem ZW, Kim JH, Loganathan GV (2001) A new heuristic optimization algorithm: harmony search. Simulation 76(2):60–68
75. Hanseth O, Aanestad M (2001) Bootstrapping networks, communities and infrastructures. On the evolution of ICT solutions in heath care. In: Proceedings of the 1st international conference on information technology in health care (ITHC'01)
76. Larrañaga P, Lozano JA (eds) (2001) Estimation of distribution algorithms: a new tool for evolutionary computation, vol 2. Springer Science & Business Media, Cham
77. Pham DT, Ghanbarzadeh A, Koç E, Otri S, Rahim S, Zaidi M (2006) The bees algorithm—a novel tool for complex optimisation problems. In: Intelligent production machines and systems, 2nd I*PROMS Virtual International Conference, pp. 454–459
78. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, vol. 200, pp. 1-10
79. Krishnanand KN, Ghose D (2006) Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications. Multiagent Grid Syst 2(3):209–222
80. Haddad OB, Afshar A, Mariño MA (2006) Honey-bees mating optimization (HBMO) algorithm: a new heuristic approach for water resources optimization. Water Resour Manag 20(5):661–680
81. Mucherino A, Seref O (2007) Monkey search: a novel metaheuristic search for global optimization. In: AIP conference proceedings, American Institute of Physics, 953(1), 162-173
82. Atashpaz-Gargari E, Lucas C (2007) Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: 2007 IEEE congress on evolutionary computation,pp. 4661–4667.
83. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12(6):702–713
84. Teodorović D (2009) Bee colony optimization (BCO). Innovations in swarm intelligence. Stud Comput Intel 248:39–60
85. He S, Wu QH, Saunders JR (2009) Group search optimizer: an optimization algorithm inspired by animal searching behavior. IEEE Trans Evol Comput 13(5):973–990
86. Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: 2009 World congress on nature & biologically inspired computing (NaBIC), pp. 210–214
87. Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inf Sci 179(13):2232–2248
88. Kashan AH (2009) League championship algorithm: a new algorithm for numerical function optimization. In: 2009 international conference of soft computing and pattern recognition, pp. 43–48.

89. Kadioglu S, Sellmann M (2009) Dialectic search. In: International conference on principles and practice of constraint programming, pp. 486–500
90. Shah-Hosseini H (2009) The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm. Int J Bio-inspired Comput 1(1–2):71–79
91. Yang XS (2009) Firefly algorithms for multimodal optimization. In: International symposium on stochastic algorithms, pp. 169–178
92. Battiti R, Brunato M, Mariello A (2019) Reactive search optimization: learning while optimizing. In: Handbook of metaheuristics, International Series in Operations Research & Management Science, vol. 272, pp. 479–511
93. Yang XS (2010) A new metaheuristic bat-inspired algorithm. In: Nature inspired cooperative strategies for optimization (NICSO 2010), studies in computational intelligence, vol. 284, pp. 65–74
94. Shah-Hosseini H (2011) Principal components analysis by the galaxy-based search algorithm: a novel metaheuristic for continuous optimisation. Int J Comput Sci Eng 6(1–2):132–140
95. Tamura K, Yasuda K (2011) Spiral dynamics inspired optimization. J Adv Comput Intell Intell Inform 15(8):1116–1122
96. Alsheddy A (2011) Empowerment scheduling: a multi-objective optimization approach using guided local search (Doctoral dissertation, University of Essex)
97. Rajabioun R (2011) Cuckoo optimization algorithm. Appl Soft Comput 11(8):5508–5518
98. Gandomi AH, Alavi AH (2012) Krill herd: a new bio-inspired optimization algorithm. Commun Nonlinear Sci Numer Simul 17(12):4831–4845
99. Civicioglu P (2012) Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm. Comput Geosci 46:229–247
100. Sadollah A, Bahreininejad A, Eskandar H, Hamdi M (2013) Mine blast algorithm: a new population based algorithm for solving constrained engineering optimization problems. Appl Soft Comput 13(5):2592–2612
101. Hatamlou A (2013) Black hole: a new heuristic optimization approach for data clustering. Inf Sci 222:175–184
102. Gandomi AH (2014) Interior search algorithm (ISA): a novel approach for global optimization. ISA Trans 53(4):1168–1183
103. Cheng MY, Prayogo D (2014) Symbiotic organisms search: a new metaheuristic optimization algorithm. Comput Struct 139:98–112
104. Kashan AH (2015) A new metaheuristic for optimization: optics inspired optimization (OIO). Comput Oper Res 55:99–125
105. Kaveh A, Mahdavi VR (2015) Colliding bodies optimization: extensions and applications. Technology & Engineering, Springer International Publishing, pp. 284
106. Salimi H (2015) Stochastic fractal search: a powerful metaheuristic algorithm. Knowl-Based Syst 75:1–18
107. Zheng YJ (2015) Water wave optimization: a new nature-inspired metaheuristic. Comput Oper Res 55:1–11
108. Doğan B, Ölmez T (2015) A new metaheuristic for numerical function optimization: Vortex search algorithm. Inf Sci 293:125–145
109. Wang GG, Deb S, Coelho LDS (2015) Elephant herding optimization. In: 2015 3rd international symposium on computational and business intelligence (ISCBI), pp. 1–5
110. Kashan AH, Akbari AA, Ostadi B (2015) Grouping evolution strategies: an effective approach for grouping problems. Appl Math Model 39(9):2703–2720
111. Mirjalili S (2016) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl 27(4):1053–1073
112. Liang YC, Cuevas Juarez JR (2016) A novel metaheuristic for continuous optimization problems: virus optimization algorithm. Eng Optim 48(1):73–93
113. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. Knowl-Based Syst 96:120–133
114. Ebrahimi A, Khamehchi E (2016) Sperm whale algorithm: an effective metaheuristic algorithm for production optimization problems. J Nat Gas Sci Eng 29:211–222
115. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191
116. Baykasoğlu A, Akpinar Ş (2017) Weighted superposition attraction (WSA): a swarm intelligence algorithm for optimization problems–Part 1: unconstrained optimization. Appl Soft Comput 56:520–540
117. Mortazavi A, Toğan V, Nuhoğlu A (2018) Interactive search algorithm: a new hybrid metaheuristic optimization algorithm. Eng Appl Artif Intell 71:275–292

118. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: algorithm and applications. Futur Gener Comput Syst 97:849–872
119. Yapici H, Cetinkaya N (2019) A new meta-heuristic optimizer: pathfinder algorithm. Appl Soft Comput 78:545–568
120. Kaur S, Awasthi LK, Sangal AL, Dhiman G (2020) Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90:103541
121. Braik M, Sheta A, Al-Hiary H (2021) A novel meta-heuristic search algorithm for solving optimization problems: capuchin search algorithm. Neural Comput Appl 33(7):2515–2547
122. Talatahari S, Azizi M, Tolouei M, Talatahari B, Sareh P (2021) Crystal structure algorithm (CryStAl): a metaheuristic optimization method. IEEE Access 9:71244–71261
123. Eslami N, Yazdani S, Mirzaei M, Hadavandi E (2022) Aphid-ant mutualism: a novel nature-inspired metaheuristic algorithm for solving optimization problems. Math Comput Simul 201:362–395
124. Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W (2022) Honey badger algorithm: new metaheuristic algorithm for solving optimization problems. Math Comput Simul 192:84–110
125. Oszust M, Sroka G, Cymerys K (2021) A hybridization approach with predicted solution candidates for improving population-based optimization algorithms. Inf Sci 574:133–161
126. Połap D, Kęsik K, Woźniak M, Damaševičius R (2018) Parallel technique for the metaheuristic algorithms using devoted local search and manipulating the solutions space. Appl Sci 8(2):293
127. Chunkai Z, Yu L, Huihe S (2000) A new evolved artificial neural network and its application. In: Proceedings of the 3rd world congress on intelligent control and automation (Cat. No. 00EX393), vol. 2, pp. 1065–1068
128. Li K, Thompson S, Wieringa PA, Peng J, Duan GR (2003) Neural networks and genetic algorithms can support human supervisory control to reduce fossil fuel power plant emissions. Cognit Technol Work 5(2):107–126
129. Leung FHF, Lam HK, Ling SH, Tam PKS (2003) Tuning of the structure and parameters of a neural network using an improved genetic algorithm. IEEE Trans Neural Netw 14(1):79–88
130. Meissner M, Schmuker M, Schneider G (2006) Optimized particle swarm optimization (OPSO) and its application to artificial neural network training. BMC Bioinform 7(1):125
131. Geethanjali M, Slochanal SMR, Bhavani R (2008) PSO trained ANN-based differential protection scheme for power transformers. Neurocomputing 71(4–6):904–918
132. Yu J, Wang S, Xi L (2008) Evolving artificial neural networks using an improved PSO and DPSO. Neurocomputing 71(4–6):1054–1060
133. Khayat O, Ebadzadeh MM, Shahdoosti HR, Rajaei R, Khajehnasiri I (2009) A novel hybrid algorithm for creating self-organizing fuzzy neural networks. Neurocomputing 73(1–3):517–524
134. Lin CJ, Hsieh MH (2009) Classification of mental task from EEG data using neural networks based on particle swarm optimization. Neurocomputing 72(4–6):1121–1130
135. Cruz-Ramírez M, Sánchez-Monedero J, Fernández-Navarro F, Fernández JC, Hervás-Martínez C (2010) Memetic pareto differential evolutionary artificial neural networks to determine growth multi-classes in predictive microbiology. Evol Intell 3(3–4):187–199
136. Malviya R, Pratihar DK (2011) Tuning of neural networks using particle swarm optimization to model MIG welding process. Swarm Evol Comput 1(4):223–235
137. Zhao L, Qian F (2011) Tuning the structure and parameters of a neural network using cooperative binary-real particle swarm optimization. Expert Syst Appl 38(5):4972–4977
138. Green RC II, Wang L, Alam M (2012) Training neural networks using central force optimization and particle swarm optimization: insights and comparisons. Expert Syst Appl 39(1):555–563
139. Vasumathi B, Moorthi S (2012) Implementation of hybrid ANN–PSO algorithm on FPGA for harmonic estimation. Eng Appl Artif Intell 25(3):476–483
140. Yaghini M, Khoshraftar MM, Fallahi M (2013) A hybrid algorithm for artificial neural network training. Eng Appl Artif Intell 26(1):293–301
141. Dragoi EN, Curteanu S, Galaction AI, Cascaval D (2013) Optimization methodology based on neural networks and self-adaptive differential evolution algorithm applied to an aerobic fermentation process. Appl Soft Comput 13(1):222–238
142. Ismail A, Jeng DS, Zhang LL (2013) An optimised product-unit neural network with a novel PSO–BP hybrid training algorithm: applications to load–deformation analysis of axially loaded piles. Eng Appl Artif Intell 26(10):2305–2314
143. Das G, Pattnaik PK, Padhy SK (2014) Artificial neural network trained by particle swarm optimization for non-linear channel equalization. Expert Syst Appl 41(7):3491–3496
144. Mirjalili S, Mirjalili SM, Lewis A (2014) Let a biogeography-based optimizer train your multi-layer perceptron. Inf Sci 269:188–209

145. Jaddi NS, Abdullah S, Hamdan AR (2015) Multi-population cooperative bat algorithm-based optimization of artificial neural network model. Inf Sci 294:628–644
146. Jaddi NS, Abdullah S, Hamdan AR (2015) Optimization of neural network model using modified bat-inspired algorithm. Appl Soft Comput 37:71–86
147. González B, Valdez F, Melin P, Prado-Arechiga G (2015) Fuzzy logic in the gravitational search algorithm enhanced using fuzzy logic with dynamic alpha parameter value adaptation for the optimization of modular neural networks in echocardiogram recognition. Appl Soft Comput 37:245–254
148. Gaxiola F, Melin P, Valdez F, Castro JR, Castillo O (2016) Optimization of type-2 fuzzy weights in backpropagation learning for neural networks using GAs and PSO. Appl Soft Comput 38:860–871
149. Karaboga D, Kaya E (2016) An adaptive and hybrid artificial bee colony algorithm (aABC) for ANFIS training. Appl Soft Comput 49:423–436
150. Jafrasteh B, Fathianpour N (2017) A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. Neurocomputing 235:217–227
151. Ganjefar S, Tofighi M (2017) Training qubit neural network with hybrid genetic algorithm and gradient descent for indirect adaptive controller design. Eng Appl Artif Intell 65:346–360
152. Aljarah I, Faris H, Mirjalili S (2018) Optimizing connection weights in neural networks using the whale optimization algorithm. Soft Comput 22(1):1–15
153. Heidari AA, Faris H, Aljarah I, Mirjalili S (2019) An efficient hybrid multilayer perceptron neural network with grasshopper optimization. Soft Comput 23(17):7941–7958
154. Hadavandi E, Mostafayi S, Soltani P (2018) A grey wolf optimizer-based neural network coupled with response surface method for modeling the strength of siro-spun yarn in spinning mills. Appl Soft Comput 72:1–13
155. Haznedar B, Kalinli A (2018) Training ANFIS structure using simulated annealing algorithm for dynamic systems identification. Neurocomputing 302:66–74
156. Pham BT, Nguyen MD, Bui KTT, Prakash I, Chapi K, Bui DT (2019) A novel artificial intelligence approach based on multi-layer perceptron neural network and biogeography-based optimization for predicting coefficient of consolidation of soil. CATENA 173:302–311
157. Han JW, Li QX, Wu HR, Zhu HJ, Song YL (2019) Prediction of cooling efficiency of forced-air pre-cooling systems based on optimized differential evolution and improved BP neural network. Appl Soft Comput 84:105733
158. Rojas-Delgado J, Trujillo-Rasúa R, Bello R (2019) A continuation approach for training Artificial Neural Networks with meta-heuristics. Pattern Recogn Lett 125:373–380
159. Khishe M, Mosavi MR (2020) Classification of underwater acoustical dataset using neural network trained by chimp optimization algorithm. Appl Acoust 157:107005
160. Wang Y, Liu H, Yu Z, Tu L (2020) An improved artificial neural network based on human-behaviour particle swarm optimization and cellular automata. Expert Syst Appl 140:112862
161. Al-Majidi SD, Abbod MF, Al-Raweshidy HS (2020) A particle swarm optimisation-trained feedforward neural network for predicting the maximum power point of a photovoltaic array. Eng Appl Artif Intell 92:103688
162. Ansari A, Ahmad IS, Bakar AA, Yaakub MR (2020) A hybrid metaheuristic method in training artificial neural network for bankruptcy prediction. IEEE Access 8:176640–176650
163. Zhang Y, Zhao J, Wang L, Wu H, Zhou R, Yu J (2021) An improved OIF Elman neural network based on CSO algorithm and its applications. Comput Commun 171:148–156
164. Li XD, Wang JS, Hao WK, Wang M, Zhang M (2022) Multi-layer perceptron classification method of medical data based on biogeography-based optimization algorithm with probability distributions. Appl Soft Comput 121:108766
165. Engel J (1988) Teaching feed-forward neural networks by simulated annealing. Complex Syst 2(6):641–648
166. Montana DJ, Davis L (1989) Training feedforward neural networks using genetic algorithms. In: IJCAI, Vol. 89, pp. 762–767
167. Whitley D, Starkweather T, Bogart C (1990) Genetic algorithms and neural networks: optimizing connections and connectivity. Parallel Comput 14(3):347–361
168. Belew RK, McInerney J, Schraudolph NN (1990) Evolving networks: using the genetic algorithm with connectionist learning. SFI studies in the sciences of complexity, pp. 511–547
169. Kitano H (1994) Neurogenetic learning: an integrated method of designing and training neural networks using genetic algorithms. Phys D Nonlinear Phenom 75(1–3):225–238
170. Battiti R, Tecchiolli G (1995) Training neural nets with the reactive tabu search. IEEE Trans Neural Netw 6(5):1185–1200

171. Yao X, Liu Y (1997) A new evolutionary system for evolving artificial neural networks. IEEE Trans Neural Netw 8(3):694–713
172. Sexton RS, Alidaee B, Dorsey RE, Johnson JD (1998) Global optimization for artificial neural networks: a tabu search application. Eur J Oper Res 106(2–3):570–584
173. Sexton RS, Dorsey RE, Johnson JD (1999) Beyond backpropagation: using simulated annealing for training neural networks. J Organ End User Comput 11(3):3–10
174. Arifovic J, Gencay R (2001) Using genetic algorithms to select architecture of a feedforward artificial neural network. Phys A Stat Mech Appl 289(3–4):574–594
175. Alvarez A (2002) A neural network with evolutionary neurons. Neural Process Lett 16(1):43–52
176. Sarkar D, Modak JM (2003) ANNSA: a hybrid artificial neural network/simulated annealing algorithm for optimal control problems. Chem Eng Sci 58(14):3131–3142
177. García-Pedrajas N, Hervás-Martínez C, Muñoz-Pérez J (2003) COVNET: a cooperative coevolutionary model for evolving artificial neural networks. IEEE Trans Neural Netw 14(3):575–596
178. Ilonen J, Kamarainen JK, Lampinen J (2003) Differential evolution training algorithm for feed-forward neural networks. Neural Process Lett 17(1):93–105
179. Augusteijn MF, Harrington TP (2004) Evolving transfer functions for artificial neural networks. Neural Comput Appl 13(1):38–46
180. Abraham A (2004) Meta learning evolutionary artificial neural networks. Neurocomputing 56:1–38
181. Lahiri A, Chakravorti S (2004) Electrode-spacer contour optimization by ANN aided genetic algorithm. IEEE Trans Dielectr Electr Insul 11(6):964–975
182. Shen Q, Jiang JH, Jiao CX, Lin WQ, Shen GL, Yu RQ (2004) Hybridized particle swarm algorithm for adaptive structure training of multilayer feed-forward neural network: QSAR studies of bioactivity of organic compounds. J Comput Chem 25(14):1726–1735
183. Kim D, Kim H, Chung D (2005) A modified genetic algorithm for fast training neural networks. In: International symposium on neural networks, pp. 660–665
184. Chatterjee A, Pulasinghe K, Watanabe K, Izumi K (2005) A particle-swarm-optimized fuzzy-neural network for voice-controlled robot systems. IEEE Trans Ind Electron 52(6):1478–1489
185. Feng P, Jie C, Xuyan T, Jiwei F (2005) Multilayered feed forward neural network based on particle swarm optimizer algorithm. J Syst Eng Electron 16(3):682–686
186. Da Y, Xiurun G (2005) An improved PSO-based ANN with simulated annealing technique. Neurocomputing 63:527–533
187. Salajegheh E, Gholizadeh S (2005) Optimum design of structures by an improved genetic algorithm using neural networks. Adv Eng Softw 36(11–12):757–767
188. Tsai JT, Chou JH, Liu TK (2006) Tuning the structure and parameters of a neural network by using hybrid Taguchi-genetic algorithm. IEEE Trans Neural Netw 17(1):69–80
189. García-Pedrajas N, Ortiz-Boyer D, Hervás-Martínez C (2006) An alternative approach for neural network evolution with a genetic algorithm: crossover by combinatorial optimization. Neural Netw 19(4):514–528
190. Ye J, Qiao J, Li MA, Ruan X (2007) A tabu based neural network learning algorithm. Neurocomputing 70(4–6):875–882
191. Socha K, Blum C (2007) An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training. Neural Comput Appl 16(3):235–247
192. Lin WQ, Jiang JH, Zhou YP, Wu HL, Shen GL, Yu RQ (2007) Support vector machine based training of multilayer feedforward neural networks as optimized by particle swarm algorithm: application in QSAR studies of bioactivity of organic compounds. J Comput Chem 28(2):519–527
193. Ulagammai M, Venkatesh P, Kannan PS, Padhy NP (2007) Application of bacterial foraging technique trained artificial and wavelet neural networks in load forecasting. Neurocomputing 70(16–18):2659–2667
194. Zhang JR, Zhang J, Lok TM, Lyu MR (2007) A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. Appl Math Comput 185(2):1026–1037
195. Lin CJ, Chen CH, Lin CT (2008) A hybrid of cooperative particle swarm optimization and cultural algorithm for neural fuzzy networks and its prediction applications. IEEE Trans Syst Man Cybern Part C (Appl Rev) 39(1):55–68
196. Tsoulos I, Gavrilis D, Glavas E (2008) Neural network construction and training using grammatical evolution. Neurocomputing 72(1–3):269–277
197. Goh CK, Teoh EJ, Tan KC (2008) Hybrid multiobjective evolutionary design for artificial neural networks. IEEE Trans Neural Netw 19(9):1531–1548
198. Bashir ZA, El-Hawary ME (2009) Applying wavelets to short-term load forecasting using PSO-based neural networks. IEEE Trans Power Syst 24(1):20–27
199. Kiranyaz S, Ince T, Yildirim A, Gabbouj M (2009) Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. Neural Netw 22(10):1448–1462

200. Slowik A (2010) Application of an adaptive differential evolution algorithm with multiple trial vectors to artificial neural network training. IEEE Trans Industr Electron 58(8):3160–3167
201. Kordík P, Koutník J, Drchal J, Kovářík O, Čepek M, Šnorek M (2010) Meta-learning approach to neural network optimization. Neural Netw 23(4):568–582
202. Lian GY, Huang KL, Chen JH, Gao FQ (2010) Training algorithm for radial basis function neural network based on quantum-behaved particle swarm optimization. Int J Comput Math 87(3):629–641
203. Zhao C, Liu X, Ding F (2010) Melt index prediction based on adaptive particle swarm optimization algorithm-optimized radial basis function neural networks. Chem Eng Technol 33(11):1909–1916
204. Ma Y, Huang M, Wan J, Hu K, Wang Y, Zhang H (2011) Hybrid artificial neural network genetic algorithm technique for modeling chemical oxygen demand removal in anoxic/oxic process. J Environ Sci Health Part A 46(6):574–580
205. Ding S, Su C, Yu J (2011) An optimizing BP neural network algorithm based on genetic algorithm. Artif Intell Rev 36(2):153–162
206. Subudhi B, Jena D (2011) A differential evolution based neural network approach to nonlinear system identification. Appl Soft Comput 11(1):861–871
207. Ghalambaz M, Noghrehabadi AR, Behrang MA, Assareh E, Ghanbarzadeh A, Hedayat N (2011) A hybrid neural network and gravitational search algorithm (HNNGSA) method to solve well known Wessinger's equation. Int J Mech Mechatron Eng 5(1):147–151
208. Irani R, Nasimi R (2011) Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir. Expert Syst Appl 38(8):9862–9866
209. Li J, Liu X (2011) Melt index prediction by RBF neural network optimized with an MPSO-SA hybrid algorithm. Neurocomputing 74(5):735–740
210. Sun J, He KY, Li H (2011) SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams. Knowl-Based Syst 24(7):1013–1023
211. Özbakır L, Delice Y (2011) Exploring comprehensible classification rules from trained neural networks integrated with a time-varying binary particle swarm optimizer. Eng Appl Artif Intell 24(3):491–500
212. Carvalho AR, Ramos FM, Chaves AA (2011) Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem. Neural Comput Appl 20(8):1273–1284
213. Han M, Fan J, Wang J (2011) A dynamic feedforward neural network based on Gaussian particle swarm optimization and its application for predictive control. IEEE Trans Neural Netw 22(9):1457–1468
214. Zanchettin C, Ludermir TB, Almeida LM (2011) Hybrid training method for MLP: optimization of architecture and training. IEEE Trans Syst Man Cybern Part B 41(4):1097–1109
215. Vadood M, Semnani D, Morshed M (2011) Optimization of acrylic dry spinning production line by using artificial neural network and genetic algorithm. J Appl Polym Sci 120(2):735–744
216. Mirjalili S, Hashim SZM, Sardroudi HM (2012) Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. Appl Math Comput 218(22):11125–11137
217. Khan K, Sahai A (2012) A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context. Int J Intell Syst Appl 4(7):23
218. Huang M, Liu X, Li J (2012) Melt index prediction by RBF neural network with an ICO-VSA hybrid optimization algorithm. J Appl Polym Sci 126(2):519–526
219. Irani R, Nasimi R (2012) An evolving neural network using an ant colony algorithm for a permeability estimation of the reservoir. Pet Sci Technol 30(4):375–384
220. Kulluk S, Ozbakir L, Baykasoglu A (2012) Training neural networks with harmony search algorithms for classification problems. Eng Appl Artif Intell 25(1):11–19
221. Nandy S, Sarkar PP, Das A (2012) Analysis of a nature inspired firefly algorithm based back-propagation neural network training. arXiv preprint arXiv:1206.5360
222. Han F, Zhu JS (2013) Improved particle swarm optimization combined with backpropagation for feedforward neural networks. Int J Intell Syst 28(3):271–288
223. Sharma N, Arun N, Ravi V (2013) An ant colony optimisation and Nelder-Mead simplex hybrid algorithm for training neural networks: an application to bankruptcy prediction in banks. Int J Inf Decis Sci 5(2):188–203
224. Li HZ, Guo S, Li CJ, Sun JQ (2013) A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. Knowl-Based Syst 37:378–387
225. Wang M, Yan X, Shi H (2013) Spatiotemporal prediction for nonlinear parabolic distributed parameter system using an artificial neural network trained by group search optimization. Neurocomputing 113:234–240
226. Lu TC, Yu GR, Juang JC (2013) Quantum-based algorithm for optimizing artificial neural networks. IEEE Trans Neural Netw Learn Syst 24(8):1266–1278
227. Askarzadeh A, Rezazadeh A (2013) Artificial neural network training using a new efficient optimization algorithm. Appl Soft Comput 13(2):1206–1213

228. Li LK, Shao S, Yiu KFC (2013) A new optimization algorithm for single hidden layer feedforward neural networks. Appl Soft Comput 13(5):2857–2862
229. Parra J, Trujillo L, Melin P (2014) Hybrid back-propagation training with evolutionary strategies. Soft Comput 18(8):1603–1614
230. Piotrowski AP (2014) Differential evolution algorithms applied to neural network training suffer from stagnation. Appl Soft Comput 21:382–406
231. Nasimi R, Irani R (2014) Identification and modeling of a yeast fermentation bioreactor using hybrid particle swarm optimization-artificial neural networks. Energy Sources Part A Recovery Util Environ Eff 36(14):1604–1611
232. Tapoglou E, Trichakis IC, Dokou Z, Nikolos IK, Karatzas GP (2014) Groundwater-level forecasting under climate change scenarios using an artificial neural network trained with particle swarm optimization. Hydrol Sci J 59(6):1225–1239
233. Raja MAZ (2014) Solution of the one-dimensional Bratu equation arising in the fuel ignition model using ANN optimised with PSO and SQP. Connect Sci 26(3):195–214
234. Beheshti Z, Shamsuddin SMH, Beheshti E, Yuhaniz SS (2014) Enhancement of artificial neural network learning using centripetal accelerated particle swarm optimization for medical diseases diagnosis. Soft Comput 18(11):2253–2270
235. Ren C, An N, Wang J, Li L, Hu B, Shang D (2014) Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl-Based Syst 56:226–239
236. Svečko R, Kusić D (2015) Feedforward neural network position control of a piezoelectric actuator based on a BAT search algorithm. Expert Syst Appl 42(13):5416–5423
237. Kumaran J, Ravi G (2015) Long-term sector-wise electrical energy forecasting using artificial neural network and biogeography-based optimization. Electr Power Compon Syst 43(11):1225–1235
238. Cui H, Feng J, Guo J, Wang T (2015) A novel single multiplicative neuron model trained by an improved glowworm swarm optimization algorithm for time series prediction. Knowl-Based Syst 88:195–209
239. Chen CH, Tsai YC, Jhang RZ (2015) Approximation of the piecewise function using neural fuzzy networks with an improved artificial bee colony algorithm. J Autom Control Eng 3(6):18–21
240. Mirjalili S (2015) How effective is the Grey Wolf optimizer in training multi-layer perceptrons. Appl Intell 43(1):150–161
241. Agrawal RK, Bawane NG (2015) Multiobjective PSO based adaption of neural network topology for pixel classification in satellite imagery. Appl Soft Comput 28:217–225
242. Gharghan SK, Nordin R, Ismail M, Abd Ali J (2015) Accurate wireless sensor localization technique based on hybrid PSO-ANN algorithm for indoor and outdoor track cycling. IEEE Sens J 16(2):529–541
243. Vadood M, Johari MS, Rahai A (2015) Developing a hybrid artificial neural network-genetic algorithm model to predict resilient modulus of polypropylene/polyester fiber-reinforced asphalt concrete. J Text Inst 106(11):1239–1250
244. Yazdi MS, Rostami SL, Kolahdooz A (2016) Optimization of geometrical parameters in a specific composite lattice structure using neural networks and ABC algorithm. J Mech Sci Technol 30(4):1763–1771
245. Jia W, Zhao D, Ding L (2016) An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample. Appl Soft Comput 48:373–384
246. Leema N, Nehemiah HK, Kannan A (2016) Neural network classifier optimization using differential evolution with global information and back propagation algorithm for clinical datasets. Appl Soft Comput 49:834–844
247. Xia R, Huang X, Li M (2016) Starch foam material performance prediction based on a radial basis function artificial neural network trained by bare-bones particle swarm optimization with an adaptive disturbance factor. J Appl Polym Sci. https://doi.org/10.1002/app.44252
248. Melo H, Watada J (2016) Gaussian-PSO with fuzzy reasoning based on structural learning for training a neural network. Neurocomputing 172:405–412
249. Chidambaram B, Ravichandran M, Seshadri A, Muniyandi V (2017) Computational heat transfer analysis and genetic algorithm-artificial neural network-genetic algorithm-based multiobjective optimization of rectangular perforated plate fins. IEEE Trans Compon Packag Manuf Technol 7(2):208–216
250. Pradeepkumar D, Ravi V (2017) Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. Appl Soft Comput 58:35–52
251. Islam B, Baharudin Z, Nallagownden P (2017) Development of chaotically improved meta-heuristics and modified BP neural network-based model for electrical energy demand prediction in smart grid. Neural Comput Appl 28(1):877–891
252. Emary E, Zawbaa HM, Grosan C (2017) Experienced gray wolf optimization through reinforcement learning and neural networks. IEEE Trans Neural Netw Learn Syst 29(3):681–694

253. Taheri K, Hasanipanah M, Golzar SB, Abd Majid MZ (2017) A hybrid artificial bee colony algorithm-artificial neural network for forecasting the blast-produced ground vibration. Eng Comput 33(3):689–700
254. Chatterjee S, Sarkar S, Hore S, Dey N, Ashour AS, Balas VE (2017) Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Comput Appl 28(8):2005–2016
255. Song LK, Fei CW, Bai GC, Yu LC (2017) Dynamic neural network method-based improved PSO and BR algorithms for transient probabilistic analysis of flexible mechanism. Adv Eng Inform 33:144–153
256. Yan D, Zhou Q, Wang J, Zhang N (2017) Bayesian regularisation neural network based on artificial intelligence optimisation. Int J Prod Res 55(8):2266–2287
257. Mansouri A, Majidi B, Shamisa A (2018) Metaheuristic neural networks for anomaly recognition in industrial sensor networks with packet latency and jitter for smart infrastructures. Int J Comput Appl 43:257–266
258. Rukhaiyar S, Alam MN, Samadhiya NK (2018) A PSO-ANN hybrid model for predicting factor of safety of slope. Int J Geotech Eng 12(6):556–566
259. Semero YK, Zhang J, Zheng D, Wei D (2018) A GA-PSO hybrid algorithm based neural network modeling technique for short-term wind power forecasting. Distrib Gener Altern Energy J 33(4):26–43
260. Bohat VK, Arya KV (2018) An effective gbest-guided gravitational search algorithm for real-parameter optimization and its application in training of feedforward neural networks. Knowl-Based Syst 143:192–207
261. Mostafaeipour A, Goli A, Qolipour M (2018) Prediction of air travel demand using a hybrid artificial neural network (ANN) with bat and firefly algorithms: a case study. J Supercomput 74(10):5461–5484
262. Camci E, Kripalani DR, Ma L, Kayacan E, Khanesar MA (2018) An aerial robot for rice farm quality inspection with type-2 fuzzy neural networks tuned by particle swarm optimization-sliding mode control hybrid algorithm. Swarm Evol Comput 41:1–8
263. Huang Y, Liu H (2018) Research on price forecasting method of China's carbon trading market based on PSO-RBF algorithm. In: International conference on bio-inspired computing: theories and applications, pp. 1–11
264. Nayak SC, Misra BB (2018) Estimating stock closing indices using a GA-weighted condensed polynomial neural network. Financ Innov 4(1):21
265. Agrawal S, Agrawal J, Kaur S, Sharma S (2018) A comparative study of fuzzy PSO and fuzzy SVD-based RBF neural network for multi-label classification. Neural Comput Appl 29(1):245–256
266. Mao WL, Hung CW (2018) Type-2 fuzzy neural network using grey wolf optimizer learning algorithm for nonlinear system identification. Microsyst Technol 24(10):4075–4088
267. Tian D, Deng J, Vinod G, Santhosh TV, Tawfik H (2018) A constraint-based genetic algorithm for optimizing neural network architectures for detection of loss of coolant accidents of nuclear power plants. Neurocomputing 322:102–119
268. Tang R, Fong S, Deb S, Vasilakos AV, Millham RC (2018) Dynamic group optimisation algorithm for training feed-forward neural networks. Neurocomputing 314:1–19
269. Xu F, Pun CM, Li H, Zhang Y, Song Y, Gao H (2019) Training feed-forward artificial neural networks with a modified artificial bee colony algorithm. Neurocomputing. https://doi.org/10.1016/j.neucom.2019.04.086
270. Karkheiran S, Kabiri-Samani A, Zekri M, Azamathulla HM (2019) Scour at bridge piers in uniform and armored beds under steady and unsteady flow conditions using ANN-APSO and ANN-GA algorithms. ISH J Hydraul Eng 27:220–228
271. Ong P, Zainuddin Z (2019) Optimizing wavelet neural networks using modified cuckoo search for multi-step ahead chaotic time series prediction. Appl Soft Comput 80:374–386
272. Harandizadeh H, Armaghani DJ, Khari M (2019) A new development of ANFIS–GMDH optimized by PSO to predict pile bearing capacity based on experimental datasets. Eng Comput 37:685–700
273. Jiang Q, Huang R, Huang Y, Chen S, He Y, Lan L, Liu C (2019) Application of BP neural network based on genetic algorithm optimization in evaluation of power grid investment risk. IEEE Access 7:154827–154835
274. Xu L, Wang H, Lin W, Gulliver TA, Le KN (2019) GWO-BP neural network based OP performance prediction for mobile multiuser communication networks. IEEE Access 7:152690–152700
275. Djema MA, Boudour M, Agbossou K, Cardenas A, Doumbia ML (2019) Adaptive direct power control based on ANN-GWO for grid interactive renewable energy systems with an improved synchronization technique. Int Trans Electr Energy Syst 29(3):e2766
276. Li A, Yang X, Xie Z, Yang C (2019) An optimized GRNN-enabled approach for power transformer fault diagnosis. IEEJ Trans Electr Electron Eng 14(8):1181–1188

277. Zhao R, Wang Y, Hu P, Jelodar H, Yuan C, Li Y, Rabbani M (2019) Selfish herds optimization algorithm with orthogonal design and information update for training multi-layer perceptron neural network. Appl Intell 49(6):2339–2381
278. Faris H, Mirjalili S, Aljarah I (2019) Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme. Int J Mach Learn Cybern 10(10):2901–2920
279. Bui QT (2019) Metaheuristic algorithms in optimizing neural network: a comparative study for forest fire susceptibility mapping in Dak Nong, Vietnam. Geomat Nat Hazards Risk 10(1):136–150
280. Yu W, Zhao F (2019) Prediction of critical properties of biodiesel fuels from FAMEs compositions using intelligent genetic algorithm-based back propagation neural network. Energy Sources Part A Recovery Util Environ Eff 43:2063–2076
281. Ma T, Wang C, Wang J, Cheng J, Chen X (2019) Particle-swarm optimization of ensemble neural networks with negative correlation learning for forecasting short-term wind speed of wind farms in western China. Inf Sci 505:157–182
282. Raval PD, Pandya AS (2020) A hybrid PSO-ANN-based fault classification system for EHV transmission lines. IETE J Res 68:3086–3099
283. Kuntoji G, Rao M, Rao S (2020) Prediction of wave transmission over submerged reef of tandem breakwater using PSO-SVM and PSO-ANN techniques. ISH J Hydraul Eng 26(3):283–290
284. da Silva Veloso YM, de Almeida MM, de Alsina OLS, Passos ML, Mujumdar AS, Leite MS (2020) Hybrid phenomenological/ANN-PSO modelling of a deformable material in spouted bed drying process. Powder Technol 366:185–196
285. Yadav A, Satyannarayana P (2020) Multi-objective genetic algorithm optimization of artificial neural network for estimating suspended sediment yield in Mahanadi River basin, India. Int J River Basin Manag 18(2):207–215
286. Wu S, Yang J, Zhang R, Ono H (2020) Prediction of endpoint sulfur content in KR desulfurization based on the hybrid algorithm combining artificial neural network with SAPSO. IEEE Access 8:33778–33791
287. Shen T, Chang J, Liang Z (2020) Swarm optimization improved BP algorithm for microchannel resistance factor. IEEE Access 8:52749–52758
288. Huang Y, Xiang Y, Zhao R, Cheng Z (2020) Air quality prediction using improved PSO-BP neural network. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2998145
289. Shen X, Zheng Y, Zhang R (2020) A hybrid forecasting model for the velocity of hybrid robotic fish based on back-propagation neural network with genetic algorithm optimization. IEEE Access 8:111731–111741
290. Ghanem WAH, Jantan A, Ghaleb SAA, Nasser AB (2020) An efficient intrusion detection model based on hybridization of artificial bee colony and dragonfly algorithms for training multilayer perceptrons. IEEE Access 8:130452–130475
291. Gong S, Gao W, Abza F (2020) Brain tumor diagnosis based on artificial neural network and a chaos whale optimization algorithm. Comput Intell 36(1):259–275
292. Zeng XP, Luo Q, Zheng JL, Chen GH (2020) An efficient neural network optimized by fruit fly optimization algorithm for user equipment association in software-defined wireless sensor network. Int J Netw Manag 30(6):e2135
293. Supraja P, Babu S, Gayathri VM, Divya G (2020) Hybrid genetic and shuffled frog-leaping algorithm for neural network structure optimization and learning model to predict free spectrum in cognitive radio. Int J Commun Syst 34:e4532
294. Fang H, Fan H, Lin S, Qing Z, Sheykhahmad FR (2020) Automatic breast cancer detection based on optimized neural network using whale optimization algorithm. Int J Imaging Syst Technol 31:425–438
295. Zafar S, Nazir M, Sabah A, Jurcut AD (2021) Securing bio-cyber interface for the internet of bio-nano things using particle swarm optimization and artificial neural networks based parameter profiling. Comput Biol Med 136:104707
296. Darabi H, Haghighi AT, Rahmati O, Shahrood AJ, Rouzbeh S, Pradhan B, Bui DT (2021) A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. J Hydrol 603:126854
297. Qiao W, Khishe M, Ravakhah S (2021) Underwater targets classification using local wavelet acoustic pattern and multi-layer perceptron neural network optimized by modified Whale optimization algorithm. Ocean Eng 219:108415
298. Zheng X, Nguyen H, Bui XN (2021) Exploring the relation between production factors, ore grades, and life of mine for forecasting mining capital cost through a novel cascade forward neural network-based salp swarm optimization model. Resour Policy 74:102300

299. Bahiraei M, Foong LK, Hosseini S, Mazaheri N (2021) Predicting heat transfer rate of a ribbed triple-tube heat exchanger working with nanofluid using neural network enhanced by advanced optimization algorithms. Powder Technol 381:459–476
300. Njock PGA, Shen SL, Zhou A, Modoni G (2021) Artificial neural network optimized by differential evolution for predicting diameters of jet grouted columns. J Rock Mech Geotech Eng 13(6):1500–1512
301. Khatir S, Tiachacht S, Le Thanh C, Ghandourah E, Mirjalili S, Wahab MA (2021) An improved Artificial Neural Network using Arithmetic Optimization Algorithm for damage assessment in FGM composite plates. Compos Struct 273:114287
302. Yeganeh A, Shadman A (2021) Using evolutionary artificial neural networks in monitoring binary and polytomous logistic profiles. J Manuf Syst 61:546–561
303. Guo Y, Yang Z, Liu K, Zhang Y, Feng W (2021) A compact and optimized neural network approach for battery state-of-charge estimation of energy storage system. Energy 219:119529
304. Korouzhdeh T, Eskandari-Naddaf H, Kazemi R (2021) Hybrid artificial neural network with biogeography-based optimization to assess the role of cement fineness on ecological footprint and mechanical properties of cement mortar expose to freezing/thawing. Constr Build Mater 304:124589
305. Li B, Ding J, Yin Z, Li K, Zhao X, Zhang L (2021) Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting. Expert Syst Appl 168:114232
306. Cui L, Tao Y, Deng J, Liu X, Xu D, Tang G (2021) BBO-BPNN and AMPSO-BPNN for multiple-criteria inventory classification. Expert Syst Appl 175:114842
307. Bai B, Zhang J, Wu X, wei Zhu G, Li X (2021) Reliability prediction-based improved dynamic weight particle swarm optimization and back propagation neural network in engineering systems. Expert Syst Appl 177:114952
308. Ghersi DE, Loubar K, Amoura M, Tazerout M (2021) Multi-objective optimization of micro co-generation spark-ignition engine fueled by biogas with various CH4/CO2 content based on GA-ANN and decision-making approaches. J Clean Prod 329:129739
309. Luo Q, Li J, Zhou Y, Liao L (2021) Using spotted hyena optimizer for training feedforward neural networks. Cogn Syst Res 65:1–16
310. Fetimi A, Dâas A, Benguerba Y, Merouani S, Hamachi M, Kebiche-Senhadji O, Hamdaoui O (2021) Optimization and prediction of safranin-O cationic dye removal from aqueous solution by emulsion liquid membrane (ELM) using artificial neural network-particle swarm optimization (ANN-PSO) hybrid model and response surface methodology (RSM). J Environ Chem Eng 9(5):105837
311. Yibre AM, Koçer B (2021) Semen quality predictive model using feed forwarded neural network trained by learning-based artificial algae algorithm. Eng Sci Technol Int J 24(2):310–318
312. Sun K, Zhao T, Li Z, Wang L, Wang R, Chen X, Yang Q, Ramezani E (2021) Methodology for optimal parametrization of the polymer membrane fuel cell based on Elman neural network method and quantum water strider algorithm. Energy Rep 7:2625–2634
313. Sheelwant A, Jadhav PM, Narala SKR (2021) ANN-GA based parametric optimization of Al-TiB2 metal matrix composite material processing technique. Mater Today Commun 27:102444
314. Medi B, Asadbeigi A (2021) Application of a GA-Optimized NNARX controller to nonlinear chemical and biochemical processes. Heliyon 7(8):e07846
315. Zhang P, Cui Z, Wang Y, Ding S (2022) Application of BPNN optimized by chaotic adaptive gravity search and particle swarm optimization algorithms for fault diagnosis of electrical machine drive system. Electr Eng 104(2):819–831
316. Zhao J, Nguyen H, Nguyen-Thoi T, Asteris PG, Zhou J (2021) Improved Levenberg–Marquardt back-propagation neural network by particle swarm and whale optimization algorithms to predict the deflection of RC beams. Eng Comput. https://doi.org/10.1007/s00366-020-01267-6
317. García-Ródenas R, Linares LJ, López-Gómez JA (2021) Memetic algorithms for training feedforward neural networks: an approach based on gravitational search algorithm. Neural Comput Appl 33(7):2561–2588
318. Uzlu E (2021) Estimates of greenhouse gas emission in Turkey with grey wolf optimizer algorithm-optimized artificial neural networks. Neural Comput Appl 33(20):13567–13585
319. Saffari A, Khishe M, Zahiri, SH (2022) Fuzzy-ChOA: an improved chimp optimization algorithm for marine mammal classification using artificial neural network. Anal Integr Circuits Signal Process 111(3):403–417
320. Liu XH, Zhang D, Zhang J, Zhang T, Zhu H (2021) A path planning method based on the particle swarm optimization trained fuzzy neural network algorithm. Clust Comput 24(3):1901–1915
321. Bui XN, Nguyen H, Tran QH, Nguyen DA, Bui HB (2021) Predicting ground vibrations due to mine blasting using a novel artificial neural network-based cuckoo search optimization. Nat Resour Res 30(3):2663–2685

322. Raei B, Ahmadi A, Neyshaburi MR, Ghorbani MA, Asadzadeh F (2021) Comparative evaluation of the whale optimization algorithm and backpropagation for training neural networks to model soil wind erodibility. Arab J Geosci 14(1):1–19
323. Cui CY, Cui W, Liu SW, Ma B (2021) An optimized neural network with a hybrid GA-ResNN training algorithm: applications in foundation pit. Arab J Geosci 14(22):1–12
324. Sağ T, Jalil AJ, Z. (2021) Vortex search optimization algorithm for training of feed-forward neural network. Int J Mach Learn Cybern 12(5):1517–1544
325. Wang T, Wang JB, Zhang XJ, Liu C (2021) A study on prediction of process parameters of shot peen forming using artificial neural network optimized by genetic algorithm. Arab J Sci Eng 46(8):7349–7361
326. Wang C, Li M, Wang R, Yu H, Wang S (2021) An image denoising method based on BP neural network optimized by improved whale optimization algorithm. EURASIP J Wirel Commun Netw 2021(1):1–22
327. Al Turki FA, Al Shammari MM (2021) Predicting the output power of a photovoltaic module using an optimized offline cascade-forward neural network-based on genetic algorithm model. Technol Econ Smart Grids Sustain Energy 6(1):1–12
328. Eappen G, Shankar T, Nilavalan R (2021) Advanced squirrel algorithm-trained neural network for efficient spectrum sensing in cognitive radio-based air traffic control application. IET Commun 15(10):1326–1351
329. Bacanin N, Bezdan T, Venkatachalam K, Zivkovic M, Strumberger I, Abouhawwash M, Ahmed AB (2021) Artificial neural networks hidden unit and weight connection optimization by quasi-refection-based learning artificial bee colony algorithm. IEEE Access 9:169135–169155
330. Liu J, Huang J, Sun R, Yu H, Xiao R (2020) Data fusion for multi-source sensors using GA-PSO-BP neural network. IEEE Trans Intell Transp Syst 22(10):6583–6598
331. Nguyen HX, Cao HQ, Nguyen TT, Tran TNC, Tran HN, Jeon JW (2021) Improving robot precision positioning using a neural network based on Levenberg Marquardt–APSO algorithm. IEEE Access 9:75415–75425
332. Ge L, Xian Y, Wang Z, Gao B, Chi F, Sun K (2020) Short-term load forecasting of regional distribution network based on generalized regression neural network optimized by grey wolf optimization algorithm. CSEE J Power Energy Syst 7(5):1093–1101
333. Kaur S, Chahal KK (2021) Prediction of Chikungunya disease using PSO-based adaptive neuro-fuzzy inference system model. Int J Comput Appl 44:641–649
334. Zhang L, Gao T, Cai G, Hai KL (2022) Research on electric vehicle charging safety warning model based on back propagation neural network optimized by improved gray wolf algorithm. J Energy Storage 49:104092
335. Guo Z, Zhang L, Chen Q, Han M, Liu W (2022) Monophenolase assay using excitation-emission matrix fluorescence and ELMAN neural network assisted by whale optimization algorithm. Anal Biochem 655:114838
336. Xue Y, Tong Y, Neri F (2022) An ensemble of differential evolution and Adam for training feed-forward neural networks. Inf Sci 608:453–471
337. Ding Z, Li J, Hao H (2022) Simultaneous identification of structural damage and nonlinear hysteresis parameters by an evolutionary algorithm-based artificial neural network. Int J Non-Linear Mech 142:103970
338. Zhu K, Shi H, Han M, Cao F (2022) Layout study of wave energy converter arrays by an artificial neural network and adaptive genetic algorithm. Ocean Eng 260:112072
339. Jnr EON, Ziggah YY, Rodrigues MJ, Relvas S (2022) A hybrid chaotic-based discrete wavelet transform and Aquila optimisation tuned-artificial neural network approach for wind speed prediction. Results Eng 14:100399
340. Zhao Y, Hu H, Song C, Wang Z (2022) Predicting compressive strength of manufactured-sand concrete using conventional and metaheuristic-tuned artificial neural network. Measurement 194:110993
341. Wu C, Wang C, Kim JW (2022) Welding sequence optimization to reduce welding distortion based on coupled artificial neural network and swarm intelligence algorithm. Eng Appl Artif Intell 114:105142
342. Si T, Bagchi J, Miranda PB (2022) Artificial neural network training using metaheuristics for medical data classification: an experimental study. Expert Syst Appl 193:116423
343. Khan A, Bukhari J, Bangash JI, Khan A, Imran M, Asim M, Khan A (2020) Optimizing connection weights of functional link neural network using APSO algorithm for medical data classification. J King Saud Univ-Comput Inf Sci 34(6):2551–2561
344. Gülcü Ş (2022) Training of the feed forward artificial neural networks using dragonfly algorithm. Appl Soft Comput 124:109023
345. Netsanet S, Zheng D, Zhang W, Teshager G (2022) Short-term PV power forecasting using variational mode decomposition integrated with Ant colony optimization and neural network. Energy Rep 8:2022–2035

346. Liang R, Le-Hung T, Nguyen-Thoi T (2022) Energy consumption prediction of air-conditioning systems in eco-buildings using hunger games search optimization-based artificial neural network model. J Build Eng 59:105087
347. Chondrodima E, Georgiou H, Pelekis N, Theodoridis Y (2022) Particle swarm optimization and RBF neural networks for public transport arrival time prediction using GTFS data. Int J Inf Manag Data Insights 2(2):100086
348. Ehteram M, Panahi F, Ahmed AN, Huang YF, Kumar P, Elshafie A (2022) Predicting evaporation with optimized artificial neural network using multi-objective salp swarm algorithm. Environ Sci Pollut Res 29(7):10675–10701
349. Li Z, Zhu B, Dai Y, Zhu W, Wang Q, Wang B (2022) Thermal error modeling of motorized spindle based on Elman neural network optimized by sparrow search algorithm. Int J Adv Manuf Technol 121:349–366
350. Ibad T, Abdulkadir SJ, Aziz N, Ragab MG, Al-Tashi Q (2022) Hyperparameter optimization of evolving spiking neural network for time-series classification. N Gener Comput 40(1):377–397
351. Foong LK, Moayedi H (2022) Slope stability evaluation using neural network optimized by equilibrium optimization and vortex search algorithm. Eng Comput 38(2):1269–1283
352. Chatterjee R, Mukherjee R, Roy PK, Pradhan DK (2022) Chaotic oppositional-based whale optimization to train a feed forward neural network. Soft Comput. https://doi.org/10.1007/s00500-022-07141-5
353. He Z, Nguyen H, Vu TH, Zhou J, Asteris PG, Mammou A (2022) Novel integrated approaches for predicting the compressibility of clay using cascade forward neural networks optimized by swarm-and evolution-based algorithms. Acta Geotech 17(4):1257–1272
354. Gülcü Ş (2021) An improved animal migration optimization algorithm to train the feed-forward artificial neural networks. Arab J Sci Eng 47:9557–9581
355. Liu G, Miao J, Zhao X, Wang Z, Li X (2022) Life prediction of residual current circuit breaker with overcurrent protection based on BP neural network optimized by genetic algorithm. J Electr Eng Technol 17(3):2003–2014
356. Al Bataineh A, Kaur D, Jalali SMJ (2022) Multi-layer perceptron training optimization using nature inspired computing. IEEE Access 10:36963–36977
357. Han HG, Sun C, Wu X, Yang H, Qiao J (2021) Training fuzzy neural network via multi-objective optimization for nonlinear systems identification. IEEE Trans Fuzzy Syst 30:3574–3588
358. Deepika D, Balaji N (2022) Effective heart disease prediction with Grey-wolf with Firefly algorithm-differential evolution (GF-DE) for feature selection and weighted ANN classification. Comput Methods Biomech Biomed Eng. https://doi.org/10.1080/10255842.2022.2078966
359. Kirankaya C, Aykut LG (2022) Training of artificial neural networks with the multi-population based artifical bee colony algorithm. Netw Comput Neural Syst 33(1):124–142
360. Yan Z, Zhu X, Wang X, Ye Z, Guo F, Xie L, Zhang G (2022) A multi-energy load prediction of a building using the multi-layer perceptron neural network method with different optimization algorithms. Energy Explor Exploit 40(4):1101–1312
361. Li Z, Piao W, Wang L, Wang X, Fu R, Fang Y (2022) China coastal bulk (Coal) freight index forecasting based on an integrated model combining ARMA, GM and BP model optimized by GA. Electronics 11(17):2732
362. Kuo CL, Kuruoglu EE, Chan WKV (2022) Neural network structure optimization by simulated annealing. Entropy 24(3):348
363. Zhao G, Wang M, Liang W (2022) A comparative study of SSA-BPNN, SSA-ENN, and SSA-SVR models for predicting the thickness of an excavation damaged zone around the roadway in rock. Mathematics 10(8):1351
364. Davar S, Nobahar M, Khan MS, Amini F (2022) The development of PSO-ANN and BOA-ANN models for predicting matric suction in expansive clay soil. Mathematics 10(16):2825
365. Huang L, Jiang L, Zhao L, Ding X (2022) Temperature compensation method based on an improved firefly algorithm optimized backpropagation neural network for micromachined silicon resonant accelerometers. Micromachines 13(7):1054
366. Wang G, Feng D, Tang W (2022) Electrical impedance tomography based on grey wolf optimized radial basis function neural network. Micromachines 13(7):1120
367. Ku KWC, Mak MW, Siu WC (1999) Adding learning to cellular genetic algorithms for training recurrent neural networks. IEEE Trans Neural Netw 10(2):239–252
368. Blanco A, Delgado M, Pegalajar MC (2001) A real-coded genetic algorithm for training recurrent neural networks. Neural Netw 14(1):93–105
369. Delgado M, Cuellar MP, Pegalajar MC (2008) Multiobjective hybrid optimization and training of recurrent neural networks. IEEE Trans Syst Man Cybern Part B (Cybern) 38(2):381–403
370. Bayer J, Wierstra D, Togelius J, Schmidhuber J (2009) Evolving memory cell structures for sequence learning. In: International conference on artificial neural networks, pp. 755–764

371. Lin CJ, Lee CY (2010) Non-linear system control using a recurrent fuzzy neural network based on improved particle swarm optimisation. Int J Syst Sci 41(4):381–395
372. Subrahmanya N, Shin YC (2010) Constructive training of recurrent neural networks using hybrid optimization. Neurocomputing 73(13–15):2624–2631
373. Hsieh TJ, Hsiao HF, Yeh WC (2011) Forecasting stock markets using wavelet transforms and recurrent neural networks: an integrated system based on artificial bee colony algorithm. Appl Soft Comput 11(2):2510–2525
374. Sheikhan M, Abbasnezhad Arabi M, Gharavian D (2015) Structure and weights optimisation of a modified Elman network emotion classifier using hybrid computational intelligence algorithms: a comparative study. Connect Sci 27(4):340–357
375. Chen S, Liu G, Wu C, Jiang Z, Chen J (2016) Image classification with stacked restricted boltzmann machines and evolutionary function array classification voter. In: 2016 IEEE congress on evolutionary computation (CEC), pp. 4599–4606
376. Real E, Moore S, Selle A, Saxena S, Suematsu YL, Tan J, Kurakin A (2017) Large-scale evolution of image classifiers. arXiv preprint arXiv:1703.01041
377. Tang X, Zhang N, Zhou J, Liu Q (2017) Hidden-layer visible deep stacking network optimized by PSO for motor imagery EEG recognition. Neurocomputing 234:1–10
378. Song Q, Zheng YJ, Xue Y, Sheng WG, Zhao MR (2017) An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. Neurocomputing 226:16–22
379. da Silva GLF, Valente TLA, Silva AC, de Paiva AC, Gattass M (2018) Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput Methods Programs Biomed 162:109–118
380. Zhou XH, Zhang MX, Xu ZG, Cai CY, Huang YJ, Zheng YJ (2019) Shallow and deep neural network training by water wave optimization. Swarm Evol Comput 50:100561
381. Shi W, Liu D, Cheng X, Li Y, Zhao Y (2019) Particle swarm optimization-based deep neural network for digital modulation recognition. IEEE Access 7:104591–104600
382. Hong YY, Taylar JV, Fajardo AC (2020) Locational marginal price forecasting using deep learning network optimized by mapping-based genetic algorithm. IEEE Access 8:91975–91988
383. Guo Y, Li JY, Zhan ZH (2020) Efficient hyperparameter optimization for convolution neural networks in deep learning: a distributed particle swarm optimization approach. Cybern Syst 52:36–57
384. ZahediNasab R, Mohseni H (2020) Neuroevolutionary based convolutional neural network with adaptive activation functions. Neurocomputing 381:306–313
385. Jallal MA, Chabaa S, Zeroual A (2020) A novel deep neural network based on randomly occurring distributed delayed PSO algorithm for monitoring the energy produced by four dual-axis solar trackers. Renew Energy 149:1182–1196
386. Elmasry W, Akbulut A, Zaim AH (2020) Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. Comput Netw 168:107042
387. Kan X, Fan Y, Fang Z, Cao L, Xiong NN, Yang D, Li X (2021) A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network. Inf Sci 568:147–162
388. Kanna PR, Santhi P (2022) Hybrid intrusion detection using mapreduce based black widow optimized convolutional long short-term memory neural networks. Expert Syst Appl 194:116545
389. Ragab M, Choudhry H, HA Asseri, Binyamin SS, Al-Rabia MW (2022) Enhanced gravitational search optimization with hybrid deep learning model for COVID-19 diagnosis on epidemiology data. In: Healthcare (Vol. 10, No. 7, p. 1339). MDPI
390. Cheung B, Sable C (2011) Hybrid evolution of convolutional networks. In: 2011 10th international conference on machine learning and applications and workshops, vol. 1, pp. 293–297
391. Desell T, Clachar S, Higgins J, Wild B (2015) Evolving deep recurrent neural networks using ant colony optimization. In: European conference on evolutionary computation in combinatorial optimization, pp. 86–98. Springer, Cham
392. Papa JP, Scheirer W, Cox DD (2016) Fine-tuning deep belief networks using harmony search. Appl Soft Comput 46:875–885
393. Zhang C, Lim P, Qin AK, Tan KC (2016) Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE Trans Neural Netw Learn Syst 28(10):2306–2318
394. Badem H, Basturk A, Caliskan A, Yuksel ME (2017) A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited–memory BFGS optimization algorithms. Neurocomputing 266:506–526
395. Gelly G, Gauvain JL (2017) Optimization of RNN-based speech activity detection. IEEE/ACM Trans Audio Speech Lang Process 26(3):646–656

396. Liu J, Gong M, Miao Q, Wang X, Li H (2017) Structure learning for deep neural networks based on multiobjective optimization. IEEE Trans Neural Netw Learn Syst 29(6):2450–2463
397. ElSaid A, Wild B, Jamiy FE, Higgins J, Desell T (2017) Optimizing LSTM RNNs using ACO to predict turbine engine vibration. In: Proceedings of the genetic and evolutionary computation conference companion, pp. 21–22
398. Kim JK, Han YS, Lee JS (2017) Particle swarm optimization–deep belief network–based rare class prediction model for highly class imbalance problem. Concurr Comput Pract Exp 29(11):e4128
399. Fujino S, Mori N, Matsumoto K (2017) Deep convolutional networks for human sketches by means of the evolutionary deep learning. In: 2017 joint 17th world congress of international fuzzy systems association and 9th international conference on soft computing and intelligent systems (IFSA-SCIS), pp. 1–5
400. Lorenzo PR, Nalepa J, Kawulok M, Ramos LS, Pastor JR (2017) Particle swarm optimization for hyper-parameter selection in deep neural networks. In: Proceedings of the genetic and evolutionary computation conference, pp. 481–488
401. Chen J, Zeng GQ, Zhou W, Du W, Lu KD (2018) Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Convers Manag 165:681–695
402. Passos LA, Rodrigues DR, Papa JP (2018) Fine tuning deep boltzmann machines through meta-heuristic approaches. In: 2018 IEEE 12th international symposium on applied computational intelligence and informatics (SACI). IEEE, pp. 000419–000424
403. Soon FC, Khaw HY, Chuah JH, Kanesan J (2018) Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition. IET Intel Transp Syst 12(8):939–946
404. ElSaid A, El Jamiy F, Higgins J, Wild B, Desell T (2018) Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration. Appl Soft Comput 73:969–991
405. Lorenzo PR, Nalepa J (2018) Memetic evolution of deep neural networks. In: Proceedings of the genetic and evolutionary computation conference, pp. 505–512
406. Pawełczyk K, Kawulok M, Nalepa J (2018) Genetically-trained deep neural networks. In: Proceedings of the genetic and evolutionary computation conference companion, pp. 63–64.
407. Fielding B, Zhang L (2018) Evolving image classification architectures with enhanced particle swarm optimisation. IEEE Access 6:68560–68575
408. Sun Y, Yen GG, Yi Z (2018) Evolving unsupervised deep neural networks for learning meaningful representations. IEEE Trans Evol Comput 23(1):89–103
409. Liang J, Meyerson E, Miikkulainen R (2018) Evolutionary architecture search for deep multitask networks. In: Proceedings of the genetic and evolutionary computation conference, pp. 466–473.
410. Khodabandehlou H, Fadali MS (2019) Training recurrent neural networks via dynamical trajectory-based optimization. Neurocomputing 368:1–10
411. Gao Y, Li Q (2019) A segmented particle swarm optimization convolutional neural network for land cover and land use classification of remote sensing images. Remote Sens Lett 10(12):1182–1191
412. Fujino S, Hatanaka T, Mori N, Matsumoto K (2019) Evolutionary deep learning based on deep convolutional neural network for anime storyboard recognition. Neurocomputing 338:393–398
413. Li Y, Xiao J, Chen Y, Jiao L (2019) Evolving deep convolutional neural networks by quantum behaved particle swarm optimization with binary encoding for image classification. Neurocomputing 362:156–165
414. Li L, Qin L, Qu X, Zhang J, Wang Y, Ran B (2019) Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. Knowl-Based Syst 172:1–14
415. Nepomuceno EG (2019) A novel method for structure selection of the recurrent random neural network using multiobjective optimisation. Appl Soft Comput 76:607–614
416. Wei P, Li Y, Zhang Z, Hu T, Li Z, Liu D (2019) An optimization method for intrusion detection classification model based on deep belief network. IEEE Access 7:87593–87605
417. Junior FEF, Yen GG (2019) Particle swarm optimization of deep neural networks architectures for image classification. Swarm Evol Comput 49:62–74
418. Navaneeth B, Suchetha M (2019) PSO optimized 1-D CNN-SVM architecture for real-time detection and classification applications. Comput Biol Med 108:85–92
419. Goel T, Murugan R, Mirjalili S, Chakrabartty DK (2020) OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. Appl Intell 51:1351–1366
420. Gao Z, Li Y, Yang Y, Wang X, Dong N, Chiang HD (2020) A GPSO-optimized convolutional neural networks for EEG-based emotion recognition. Neurocomputing 380:225–235
421. Martín A, Vargas VM, Gutiérrez PA, Camacho D, Hervás-Martínez C (2020) Optimising convolutional neural networks using a hybrid statistically-driven coral reef optimisation algorithm. Appl Soft Comput 90:106–144

422. Tang J, Zeng J, Wang Y, Yuan H, Liu F, Huang H (2020) Traffic flow prediction on urban road network based on license plate recognition data: combining attention-LSTM with genetic algorithm. Transp Transp Sci 17:1217–1243
423. Lima LL, Ferreira Junior JR, Oliveira MC (2020) Toward classifying small lung nodules with hyperparameter optimization of convolutional neural networks. Comput Intell 37:1599–1618
424. Renukadevi T, Karunakaran S (2020) Optimizing deep belief network parameters using grasshopper algorithm for liver disease classification. Int J Imaging Syst Technol 30(1):168–184
425. Ali SA, Raza B, Malik AK, Shahid AR, Faheem M, Alquhayz H, Kumar YJ (2020) An optimally configured and improved deep belief network (OCI-DBN) approach for heart disease prediction based on ruzzo-tompa and stacked genetic algorithm. IEEE Access 8:65947–65958
426. Rajagopal A, Joshi GP, Ramachandran A, Subhalakshmi RT, Khari M, Jha S, Shankar K, You J (2020) A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles. IEEE Access 8:135383–135393
427. Lu Z, Whalen I, Dhebar Y, Deb K, Goodman E, Banzhaf W, Boddeti VN (2020) Multi-objective evolutionary design of deep convolutional neural networks for image classification. IEEE Trans Evol Comput 25:277–291
428. Lin Y, Chen C, Xiao F, Avatefipour O, Alsubhi K, Yunianta A (2020) An evolutionary deep learning anomaly detection framework for in-vehicle networks-CAN bus. IEEE Trans Ind Appl. https://doi.org/10.1109/TIA.2020.3009906
429. Kavousi-Fard A, Dabbaghjamanesh M, Jin T, Su W, Roustaei M (2020) An evolutionary deep learning-based anomaly detection model for securing vehicles. IEEE Trans Intell Transp Syst 22:4478–4486
430. Johnson F, Valderrama A, Valle C, Crawford B, Soto R, Ñanculef R (2020) Automating configuration of convolutional neural network hyperparameters using genetic algorithm. IEEE Access 8:156139–156152
431. Zheng Y, Fu H, Li R, Hsung TC, Song Z, Wen D (2021) Deep neural network oriented evolutionary parametric eye modeling. Pattern Recogn 113:107755
432. Pang L, Wang L, Yuan P, Yan L, Yang Q, Xiao J (2021) Feasibility study on identifying seed viability of Sophora japonica with optimized deep neural network and hyperspectral imaging. Comput Electron Agric 190:106426
433. Gai J, Zhong K, Du X, Yan K, Shen J (2021) Detection of gear fault severity based on parameter-optimized deep belief network using sparrow search algorithm. Measurement 185:110079
434. Sun X, Wang G, Xu L, Yuan H, Yousefi N (2021) Optimal estimation of the PEM fuel cells applying deep belief network optimized by improved archimedes optimization algorithm. Energy 237:121532
435. Samir AA, Rashwan AR, Sallam KM, Chakrabortty RK, Ryan MJ, Abohany AA (2021) Evolutionary algorithm-based convolutional neural network for predicting heart diseases. Comput Ind Eng 161:107651
436. Liu D, Ding W, Dong ZS, Pedrycz W (2022) Optimizing deep neural networks to predict the effect of social distancing on COVID-19 spread. Comput Ind Eng 166:107970
437. Mao WL, Chen WC, Wang CT, Lin YH (2021) Recycling waste classification using optimized convolutional neural network. Resour Conserv Recycl 164:105132
438. Kim TY, Cho SB (2021) Optimizing CNN-LSTM neural networks with PSO for anomalous query access control. Neurocomputing 456:666–677
439. Zhang L, Lim CP, Yu Y (2021) Intelligent human action recognition using an ensemble model of evolving deep networks with swarm-based optimization. Knowl-Based Syst 220:106918
440. Li C, Yin C, Xu X (2021) Hybrid optimization assisted deep convolutional neural network for hardening prediction in steel. J King Saud Univ-Sci 33(6):101453
441. Mohakud R, Dash R (2022) Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN. J King Saud Univ-Comput Inf Sci 34:6505–7840
442. Altan A, Karasu S, Zio E (2021) A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. Appl Soft Comput 100:106996
443. Roder M, Passos LA, de Rosa GH, de Albuquerque VHC, Papa JP (2021) Reinforcing learning in deep belief networks through nature-inspired optimization. Appl Soft Comput 108:107466
444. Mathe M, Padmaja M, Krishna BT (2021) Intelligent approach for artifacts removal from EEG signal using heuristic-based convolutional neural network. Biomed Signal Process Control 70:102935
445. Mahesh DB, Murty GS, Lakshmi DR (2021) Optimized local weber and gradient pattern-based medical image retrieval and optimized convolutional neural network-based classification. Biomed Signal Process Control 70:102971
446. Singh P, Chaudhury S, Panigrahi BK (2021) Hybrid MPSO-CNN: Multi-level particle swarm optimized hyperparameters of convolutional neural network. Swarm Evol Comput 63:100863
447. Kumar K, Haider M, Uddin T (2021) Enhanced prediction of intra-day stock market using metaheuristic optimization on RNN–LSTM network. N Gener Comput 39(1):231–272

448. Kumar P, Batra S, Raman B (2021) Deep neural network hyper-parameter tuning through twofold genetic approach. Soft Comput 25(13):8747–8771
449. Chitra B, Kumar SS (2021) An optimized deep learning model using mutation-based atom search optimization algorithm for cervical cancer detection. Soft Comput 25(24):15363–15376
450. Deighan DS, Field SE, Capano CD, Khanna G (2021) Genetic-algorithm-optimized neural networks for gravitational wave classification. Neural Comput Appl 33(20):13859–13883
451. Qu J, Liu F, Ma Y (2022) A dual encoder DAE neural network for imbalanced binary classification based on NSGA-III and GAN. Pattern Anal Appl 25(1):17–34
452. Goel T, Murugan R, Mirjalili S, Chakrabartty DK (2021) OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. Appl Intell 51(3):1351–1366
453. Liu B, Nie L (2021) Gradient based invasive weed optimization algorithm for the training of deep neural network. Multimed Tools Appl 80(15):22795–22819
454. Kumar R, Kumar P, Kumar Y (2021) Integrating big data driven sentiments polarity and ABC-optimized LSTM for time series forecasting. Multimed Tools Appl. https://doi.org/10.1007/s11042-020-08904-8
455. Das D, Das AK, Pal AR, Jaypuria S, Pratihar DK, Roy GG (2021) Meta-heuristic algorithms-tuned Elman vs. Jordan recurrent neural networks for modeling of electron beam welding process. Neural Process Lett 53(2):1647–1663
456. Gong C, Wang X, Gani A, Qi H (2021) Enhanced long short-term memory with fireworks algorithm and mutation operator. J Supercomput 77(11):12630–12646
457. Chen Z, Yang C, Qiao J (2022) The optimal design and application of LSTM neural network based on the hybrid coding PSO algorithm. J Supercomput 78(5):7227–7259
458. Bacanin N, Bezdan T, Venkatachalam K, Al-Turjman F (2021) Optimized convolutional neural network by firefly algorithm for magnetic resonance image classification of glioma brain tumor grade. J Real-Time Image Proc 18(4):1085–1098
459. Akin Sherly LT, Jaya T (2021) Improved firefly algorithm-based optimized convolution neural network for scene character recognition. SIViP 15(5):885–893
460. Datta S, Chakrabarti S (2021) Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer. Sādhanā 46(2):1–23
461. Alenazy WM, Alqahtani AS (2021) Gravitational search algorithm based optimized deep learning model with diverse set of features for facial expression recognition. J Ambient Intell Humaniz Comput 12(2):1631–1646
462. Sudha MS, Valarmathi K (2021) An optimized deep belief network to detect anomalous behavior in social media. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02708-2
463. Jammalamadaka K, Parveen N (2021) Testing coverage criteria for optimized deep belief network with search and rescue. J Big Data 8(1):1–20
464. Gadekallu TR, Alazab M, Kaluri R, Maddikunta PKR, Bhattacharya S, Lakshmanna K (2021) Hand gesture classification using a novel CNN-crow search algorithm. Complex Intell Syst 7(4):1855–1868
465. Irmak E (2021) Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework. Iran J Sci Technol Trans Electr Eng 45(3):1015–1036
466. Arjunagi S, Patil NB (2021) Optimized convolutional neural network for identification of maize leaf diseases with adaptive ageist spider monkey optimization model. Int J Inf Technol. https://doi.org/10.1007/s41870-021-00657-3
467. Li P, Wang S, Ji H, Zhan Y, Li H (2021) Air quality index prediction based on an adaptive dynamic particle swarm optimized bidirectional gated recurrent neural network-china region. Adv Theory Simul 4(12):2100220
468. Oyelade ON, Ezugwu AE (2022) Characterization of abnormalities in breast cancer images using nature-inspired metaheuristic optimized convolutional neural networks model. Concurr Comput Pract Exp 34(4):e6629
469. Tripathi MK, Maktedar DD (2021) Optimized deep learning model for mango grading: hybridizing lion plus firefly algorithm. IET Image Proc 15(9):1940–1956
470. Karuppusamy L, Ravi J, Dabbu M, Lakshmanan S (2022) Chronological salp swarm algorithm based deep belief network for intrusion detection in cloud using fuzzy entropy. Int J Numer Model Electron Netw Devices Fields 35(1):e2948
471. Krishna Priya R, Chacko S (2021) Improved particle swarm optimized deep convolutional neural network with super-pixel clustering for multiple sclerosis lesion segmentation in brain MRI imaging. Int J Numer Methods Biomed Eng 37(9):e3506
472. Danesh K, Vasuhi S (2021) An effective spectrum sensing in cognitive radio networks using improved convolution neural network by glow worm swarm algorithm. Trans Emerg Telecommun Technol 32(11):1–20

473. Zhang J, Sun G, Sun Y, Dou H, Bilal A (2021) Hyper-parameter optimization by using the genetic algorithm for upper limb activities recognition based on neural networks. IEEE Sens J 21(2):1877–1884
474. Farrag TA, Elattar EE (2021) Optimized Deep stacked long short-term memory network for long-term load forecasting. IEEE Access 9:68511–68522
475. Arora P, Jalali SMJ, Ahmadian S, Panigrahi BK, Suganthan P, Khosravi A (2022) Probabilistic wind power forecasting using optimised deep auto-regressive recurrent neural networks. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2022.3160696
476. Goay CH, Ahmad NS, Goh P (2021) Transient simulations of high-speed channels using CNN-LSTM with an adaptive successive halving algorithm for automated hyperparameter optimizations. IEEE Access 9:127644–127663
477. Liu X, Shi Q, Liu Z, Yuan J (2021) Using LSTM neural network based on improved PSO and attention mechanism for predicting the effluent COD in a wastewater treatment plant. IEEE Access 9:146082–146096
478. Davoudi K, Thulasiraman P (2021) Evolving convolutional neural network parameters through the genetic algorithm for the breast cancer classification problem. Simulation 97(8):511–527
479. Liu X, Zhang C, Cai Z, Yang J, Zhou Z, Gong X (2021) Continuous particle swarm optimization-based deep learning architecture search for hyperspectral image classification. Remote Sens 13(6):1082
480. Brodzicki A, Piekarski M, Jaworek-Korjakowska J (2021) The whale optimization algorithm approach for deep neural networks. Sensors 21(23):8003
481. Baniasadi S, Rostami O, Martín D, Kaveh M (2022) A novel deep supervised learning-based approach for intrusion detection in IoT systems. Sensors 22(12):4459
482. Paul V, Ramesh R, Sreeja P, Jarin T, Kumar PS, Ansar S, Ashraf GA, Pandey S, Said Z (2022) Hybridization of long short-term memory with sparrow search optimization model for water quality index prediction. Chemosphere 307:135762
483. Gonçalves CB, Souza JR, Fernandes H (2022) CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images. Comput Biol Med 142:105205
484. Muthukannan P (2022) Optimized convolution neural network based multiple eye disease detection. Comput Biol Med 146:105648
485. Xu Y, Hu C, Wu Q, Jian S, Li Z, Chen Y, Zhang G, Zhang Z, Wang S (2022) Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. J Hydrol 608:127553
486. Antony Raj S, Giftson Samuel G (2022) BOSS-D-RBFN: BOosted Salp Swarm optimization based Deep RBFN for MPPT under partial shading condition in photovoltaic systems. Optik 259:168876
487. Hassanzadeh T, Essam D, Sarker R (2022) EvoDCNN: an evolutionary deep convolutional neural network for image classification. Neurocomputing 488:271–283
488. Palaniswamy T (2022) Hyperparameter optimization based deep convolution neural network model for automated bone age assessment and classification. Displays 73:102206
489. Jalali SMJ, Ahmadian S, Khodayar M, Khosravi A, Shafie-khah M, Nahavandi S, Catalão JP (2022) An advanced short-term wind power forecasting framework based on the optimized deep neural network models. Int J Electr Power Energy Syst 141:108143
490. Lokku G, Reddy GH, Prasad MG (2022) OPFaceNet: OPtimized Face Recognition Network for noise and occlusion affected face images using hyperparameters tuned convolutional neural network. Appl Soft Comput 117:108365
491. Ewees AA, Al-qaness MA, Abualigah L, Abd Elaziz M (2022) HBO-LSTM: optimized long short term memory with heap-based optimizer for wind power forecasting. Energy Convers Manag 268:116022
492. Huo F, Chen Y, Ren W, Dong H, Yu T, Zhang J (2022) Prediction of reservoir key parameters in 'sweet spot'on the basis of particle swarm optimization to TCN-LSTM network. J Petrol Sci Eng 214:110544
493. Li W, Wang L, Dong Z, Wang R, Qu B (2022) Reservoir production prediction with optimized artificial neural network and time series approaches. J Petrol Sci Eng 215:110586
494. Ge S, Gao W, Cui S, Chen X, Wang S (2022) Safety prediction of shield tunnel construction using deep belief network and whale optimization algorithm. Autom Constr 142:104488
495. Jalali SMJ, Ahmadian M, Ahmadian S, Hedjam R, Khosravi A, Nahavandi S (2022) X-ray image based COVID-19 detection using evolutionary deep learning approach. Expert Syst Appl 201:116942
496. Li Y, Peng T, Zhang C, Sun W, Hua L, Ji C, Shahzad NM (2022) Multi-step ahead wind speed forecasting approach coupling maximal overlap discrete wavelet transform, improved grey wolf optimization algorithm and long short-term memory. Renew Energy 196:1115–1126
497. Veluchamy S, Thirumalai J, Sureshkanna P (2022) RBorderNet: Rider Border Collie Optimization-based Deep Convolutional Neural Network for road scene segmentation and road intersection classification. Digit Signal Process 129:103626

498. Mohakud R, Dash R (2021) Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection. J King Saud Univ-Comput Inf Sci 34(8):6280–6291
499. Ahmad J, Shah SA, Latif S, Ahmed F, Zou Z, Pitropakis N (2022) DRaNN_PSO: A deep random neural network with particle swarm optimization for intrusion detection in the industrial internet of things. J King Saud Univ-Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2022.07.023
500. Chen F, Yang C, Khishe M (2022) Diagnose Parkinson's disease and cleft lip and palate using deep convolutional neural networks evolved by IP-based chimp optimization algorithm. Biomed Signal Process Control 77:103688
501. Karthiga M, Santhi V, Sountharrajan S (2022) Hybrid optimized convolutional neural network for efficient classification of ECG signals in healthcare monitoring. Biomed Signal Process Control 76:103731
502. Kanipriya M, Hemalatha C, Sridevi N, SriVidhya SR, Shabu SJ (2022) An improved capuchin search algorithm optimized hybrid CNN-LSTM architecture for malignant lung nodule detection. Biomed Signal Process Control 78:103973
503. Hu H, Xia X, Luo Y, Zhang C, Nazir MS, Peng T (2022) Development and application of an evolutionary deep learning framework of LSTM based on improved grasshopper optimization algorithm for short-term load forecasting. J Build Eng 57:104975
504. Raziani S, Azimbagirad M (2022) Deep CNN hyperparameter optimization algorithms for sensor-based human activity recognition. Neurosci Inform 2:100078
505. Falahzadeh MR, Farokhi F, Harimi A, Sabbaghi-Nadooshan R (2022) Deep convolutional neural network and gray wolf optimization algorithm for speech emotion recognition. Circuits Syst Signal Process. https://doi.org/10.1007/s00034-022-02130-3
506. Vigneshwaran B, Iruthayarajan MW, Maheswari RV (2022) Enhanced particle swarm optimization-based convolution neural network hyperparameters tuning for transformer failure diagnosis under complex data sources. Electr Eng. https://doi.org/10.1007/s00202-022-01501-y
507. Jalali SMJ, Ahmadian S, Khodayar M, Khosravi A, Ghasemi V, Shafie-khah M, Nahavandi S, Catalão JP (2021) Towards novel deep neuroevolution models: chaotic levy grasshopper optimization for short-term wind speed forecasting. Eng Comput 38:1787–1811
508. Surya V, Senthilselvi A (2022) Identification of oil authenticity and adulteration using deep long short-term memory-based neural network with seagull optimization algorithm. Neural Comput Appl 34(10):7611–7625
509. Balasubramanian K, Ananthamoorthy NP, Ramya K (2022) An approach to classify white blood cells using convolutional neural network optimized by particle swarm optimization algorithm. Neural Comput Appl. https://doi.org/10.1007/s00521-022-07279-1
510. Pandey A, Jain K (2022) Plant leaf disease classification using deep attention residual network optimized by opposition-based symbiotic organisms search algorithm. Neural Comput Appl. https://doi.org/10.1007/s00521-022-07587-6
511. Challapalli JR, Devarakonda N (2022) A novel approach for optimization of convolution neural network with hybrid particle swarm and grey wolf algorithm for classification of Indian classical dances. Knowl Inf Syst. https://doi.org/10.1007/s10115-022-01707-3
512. Rodrigues LF, Backes AR, Travençolo BAN, de Oliveira GMB (2022) Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification. J Digit Imaging 35(3):623–637
513. Sasank VVS, Venkateswarlu S (2022) Hybrid deep neural network with adaptive rain optimizer algorithm for multi-grade brain tumor classification of MRI images. Multimed Tools Appl 81(6):8021–8057
514. Kavitha TS, Prasad D, Satya K (2022) A novel method of compressive sensing MRI reconstruction based on sandpiper optimization algorithm (SPO) and mask region based convolution neural network (mask RCNN). Multimed Tools Appl. https://doi.org/10.1007/s11042-022-12940-x
515. Qader SM, Hassan BA, Rashid TA (2022) An improved deep convolutional neural network by using hybrid optimization algorithms to detect and classify brain tumor using augmented MRI images. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-13260-w
516. Karthik E, Sethukarasi T (2022) A centered convolutional restricted boltzmann machine optimized by hybrid atom search arithmetic optimization algorithm for sentimental analysis. Neural Process Lett 54:4123–4151
517. Li BJ, Sun GL, Liu Y, Wang WC, Huang XD (2022) monthly runoff forecasting using variational mode decomposition coupled with gray wolf optimizer-based long short-term memory neural networks. Water Resour Manag 36(6):2095–2115
518. Bhardwaj S, Agarwal R (2022) An efficient speaker identification framework based on Mask R-CNN classifier parameter optimized using hosted cuckoo optimization (HCO). J Ambient Intell Humaniz Comput 13:1–13

519. Kaushik A, Singal N, Prasad M (2022) Incorporating whale optimization algorithm with deep belief network for software development effort estimation. Int J Syst Assur Eng Manag 13:1637–1651
520. Liu J, Jiang R, Zhu D, Zhao J (2022) Short-term subway inbound passenger flow prediction based on AFC Data and PSO-LSTM optimized model. Urban Rail Transit 8(1):56–66
521. Souissi B, Ghorbel A (2022) Upper confidence bound integrated genetic algorithm-optimized long short-term memory network for click-through rate prediction. Appl Stoch Models Bus Ind 38(3):475–496
522. Balasubramanian K, Kishore R, Krishnamoorthy GD (2022) Optimal knee osteoarthritis diagnosis using hybrid deep belief network based on Salp swarm optimization method. Concurr Comput Pract Exp 34(13):e6913
523. Mukherjee G, Chatterjee A, Tudu B (2022) Identification of the types of disease for tomato plants using a modified gray wolf optimization optimized MobileNetV2 convolutional neural network architecture driven computer vision framework. Concurr Comput Pract Exp 34(22):e7161
524. Ponmalar A, Dhanakoti V (2022) Hybrid Whale Tabu algorithm optimized convolutional neural network architecture for intrusion detection in big data. Concurr Comput Pract Exp. https://doi.org/10.1002/cpe.7038
525. Suresh T, Brijet Z, Subha TD (2022) Modified local binary patterns based feature extraction and hyper parameters tuned attention segmental recurrent neural network classifier using flamingo search optimization algorithm for disease diagnosis model. Concurr Comput Pract Exp. https://doi.org/10.1002/cpe.7182
526. Xu X, Liu C, Zhao Y, Lv X (2022) Short-term traffic flow prediction based on whale optimization algorithm optimized BiLSTM_Attention. Concurr Comput Pract Exp 34(10):e6782
527. Tuerxun W, Xu C, Guo H, Guo L, Zeng N, Cheng Z (2022) An ultra-short-term wind speed prediction model using LSTM based on modified tuna swarm optimization and successive variational mode decomposition. Energy Sci Eng. https://doi.org/10.1002/ese3.1183
528. Chandraraju TS, Jeyaprakash A (2022) Categorization of breast masses based on deep belief network parameters optimized using chaotic krill herd optimization algorithm for frequent diagnosis of breast abnormalities. Int J Imaging Syst Technol 32:1561–1576
529. Jiang Y, Xia L, Zhang J (2021) A fault feature extraction method for DC-DC converters based on automatic hyperparameter-optimized one-dimensional convolution and long short-term memory neural networks. IEEE J Emerg Sel Top Power Elect 10(4):4703–4714
530. Fetanat M, Stevens M, Jain P, Hayward C, Meijering E, Lovell NH (2021) Fully Elman neural network: a novel deep recurrent neural network optimized by an improved harris hawks algorithm for classification of pulmonary arterial wedge pressure. IEEE Trans Biomed Eng 69(5):1733–1744
531. Jiang Y, Jia M, Zhang B, Deng L (2022) Ship attitude prediction model based on cross-parallel algorithm optimized neural network. IEEE Access 10:77857–77871
532. Gampala V, Rathan K, Shajin FH, Rajesh P (2022) Diagnosis of COVID-19 patients by adapting hyper parametertuned deep belief network using hosted cuckoo optimization algorithm. Electromagn Biol Med. https://doi.org/10.1080/15368378.2022.2065679
533. Li Q, Yang M, Lu Z, Zhang Y, Ba W (2022) A soft-sensing method for product quality monitoring based on particle swarm optimization deep belief networks. Trans Inst Meas Control. https://doi.org/10.1177/01423312221093166
534. Yu Y, Rashidi M, Samali B, Mohammadi M, Nguyen TN, Zhou X (2022) Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm. Struct Health Monit. https://doi.org/10.1177/14759217211053546
535. Li X, Li Y, Cao Y, Duan S, Wang X, Zhao Z (2022) Fault diagnosis method for aircraft EHA based on FCNN and MSPSO hyperparameter optimization. Appl Sci 12(17):8562
536. Pellegrino E, Brunet T, Pissier C, Camilla C, Abbou N, Beaufils N, Nanni-Metellus I, Métellus P, Ouafik LH (2022) Deep learning architecture optimization with metaheuristic algorithms for predicting BRCA1/BRCA2 pathogenicity NGS analysis. BioMedInformatics 2(2):244–267
537. Mohapatra M, Parida AK, Mallick PK, Zymbler M, Kumar S (2022) Botanical leaf disease detection and classification using convolutional neural network: a hybrid metaheuristic enabled approach. Computers 11(5):82
538. Shankar K, Kumar S, Dutta AK, Alkhayyat A, Jawad AJAM, Abbas AH, Yousif YK (2022) An automated hyperparameter tuning recurrent neural network model for fruit classification. Mathematics 10(13):2358
539. Fan Y, Zhang Y, Guo B, Luo X, Peng Q, Jin Z (2022) A hybrid sparrow search algorithm of the hyperparameter optimization in deep learning. Mathematics 10(16):3019
540. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1(1):67–82
541. Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Comput 8(7):1341–1390

542. Kaveh M, Mesgari MS (2019) Hospital site selection using hybrid PSO algorithm-case study: district 2 of Tehran. Sci-Res J Geogr Data 28(111):7–22
543. Kaveh M, Mesgari MS (2019) Improved biogeography-based optimization using migration process adjustment: an approach for location-allocation of ambulances. Comput Ind Eng 135:800–813
544. Reddy KK, Sarkar S, Venugopalan V, Giering M (2016) Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach. In: Annual conference of the prognostics and health management society, Vol. 2016
545. Liu X, Gao J, He X, Deng L, Duh K, Wang YY (2015) Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In: Proceedings of NAACL, pp. 912–921