Chapter 2

Background and Related Works

Recently, AI has emerged as a hot topic, presenting impressive results with EC. AI, a buzzword for today's era, involves advanced ML methods emphasizing high-level data representation learning rather than task-specific learning [24]. ML has been extensively used due to advances in hardware and the development of computational fast and efficient optimization algorithms [25]. In addition, ML can be used to evaluate and mine exciting Big Data models [26]. Like several other problems in AI, ML also consist of an optimization problem within itself that may vary in a range of complexity and type. On the other hand, the application diversity of EC methods is enormous, and the literature is growing quickly. It also shifts the research paradigm, culminating in Evolutionary ML (EML), and is regarded as a new revolution in AI.

As a result, this chapter provides a systematic review of studies published in the last ten years on EC-based frameworks for large-scale optimization and ML. There are also a few early works that are over ten years old but have made significant contributions to the field. These systematic reviews aided us in determining our objectives by identifying research gaps, challenges, and open issues in EC and other applications.

2.1 Background

Before proceeding with the related work, we examine some of the essential preliminary terms and methods.

2.1.1 Basic Definitions

Definition 2.1 ***Optimization***: *An optimization problem that consists of only one objective function that needs to be minimized or maximized is known as a single objective optimization problem. Mathematically, for a minimization problem, it is defined as :*

$$\begin{aligned} minimize_{\vec{x}}\, f(\vec{x}) \\ subject\ to \quad \vec{x} \in X \subseteq \mathbb{R}^n \end{aligned} \tag{2.1}$$

and $\vec{x}^*$ is an optimal solution in feasible set X for problem in (2.1) if it satisfies $f(\vec{x}^*) \leq f(\vec{x}) \quad \forall x \in X$

Definition 2.2 ***Multi-objective Optimization***: *It is defined as an optimization problem with several objective functions that are usually conflicting and need to be optimized simultaneously. It is also called vector optimization or multi-criteria optimization, which gives a set of equally good solutions instead of a single superior solution. Mathematically, it is formulated as follows:*

$$\begin{aligned} Minimize\, F(\vec{x}) \\ s.t\ \vec{x} \in \Omega \\ where\ F(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}) \cdots, f_p(\vec{x})]^T \end{aligned} \tag{2.2}$$

where Ω is the search space and $\vec{x}$ is a decision vector such that $F : \Omega \rightarrow \mathbb{R}^p$ with p number of objective functions in $\mathbb{R}^p$ objective space and $\Omega \subset \mathbb{R}^m$

Definition 2.3 ***Pareto Dominance***: *Given two vectors $\vec{a}$ and $\vec{b}$ such that $\vec{a} = (a_1, a_2, \cdots, a_m)$ and $\vec{b} = (b_1, b_2, \cdots, b_m)$, $\vec{a}$ is said to dominate $\vec{b}$ i.e.,($\vec{a} \prec \vec{b}$) if and only if it satisfies:*

$$\forall\, i \in \{1, 2, .., m\}, [f_i(\vec{a}) \leq f_i(\vec{b})] \wedge \exists\, \{i \in 1, 2, \cdots, p\} : [f_i(\vec{a}) < f_i(\vec{b}] \tag{2.3}$$

Definition 2.4 ***Pareto Optimality***: *Let $\mathbb{S}$ be a set of solutions for a given multi-objective optimization problem in (2.2). A solution $\vec{s} \in \mathbb{S}$ is said to be pareto optimal iff:*

$$\nexists\, \vec{s'} \in \mathbb{S} \,|\, \vec{s'} \prec \vec{s} \tag{2.4}$$

Definition 2.5 ***Pareto Optimal Set and Pareto Optimal Front***: *A Pareto optimal set refers to a set consisting of all Pareto optimal solutions which represent the best trade-offs between multiple objectives of a problem and the corresponding objective function's values make up the Pareto optimal front.*

Definition 2.6 ***Encryption and Decryption*** : *Encryption E is a process of encoding that uses a finite set of instructions known as an algorithm to transform the original image I into unreadable form CI to others except to the receiver having secret key K to retrieve the original image. The process of recovering the original image is known as decryption and is denoted by D. Mathematically it is stated as follows:*

$$CI = E(K, I) \tag{2.5}$$

$$I = D(K, CI) \tag{2.6}$$

Definition 2.7 ***Feature Selection***: *Feature selection (FS) techniques are designed to reduce the number of irrelevant, noisy, and redundant attributes in the dataset without compromising the performance of the model and an important step to overcoming the problem of "curse of dimensionality". FS problem (ρ, P) can be formulated as an* *optimization problem as identify* F^* *for which*

$$P(F^*) = min_{F \in \rho} P(F,V) \tag{2.7}$$

where ρ *is the set of all possible feature subsets and* $F \subset \rho$ *is feature subset and* $P : \rho \times \psi \rightarrow (R)$ *represents a criterion to measure the quality of* F *w.r.t's utility in classifying points set* $V \in \psi$ *. Each member of* V *is a* d *dimensional vector and is projected in* d_F *dimensional subspace such that* $d_F = \mod F \leq d$ *defined by* F*.*

2.1.2 Evolutionary Computation: An Overview

Evolutionary Computation (EC) is a subfield of AI. The mechanisms of biological evolution and natural events performed by organisms serve as inspiration for designing EC methods. Generally, EC [27] is defined as a computing tool that follows the Darwinian evolution principle to solve an optimization problem. It sought to develop novel

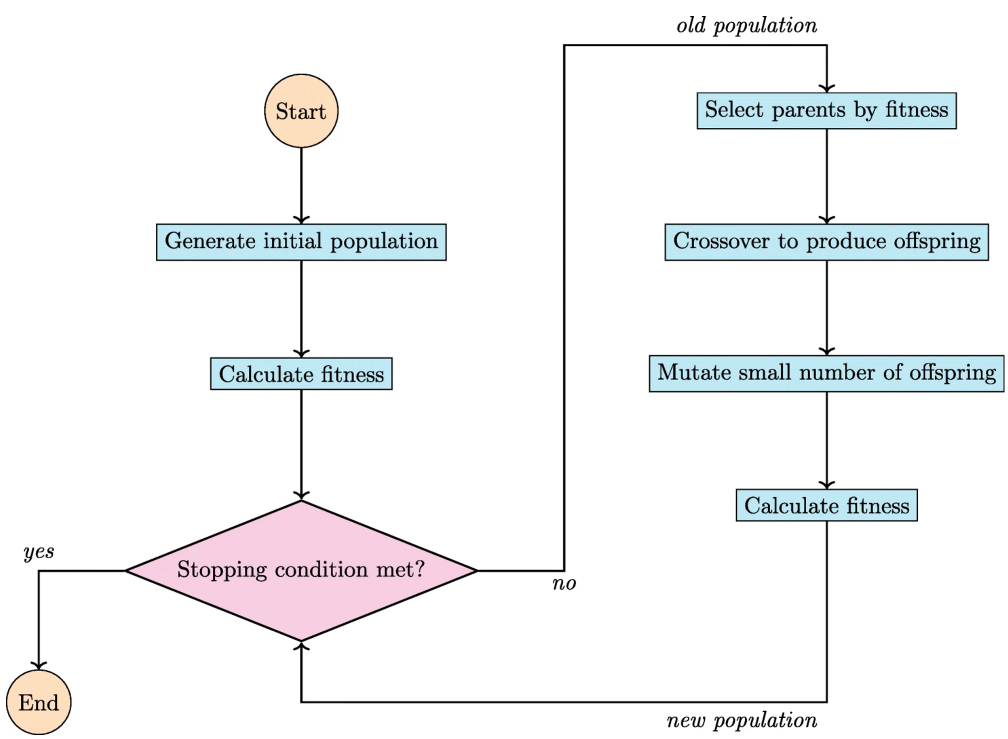

Figure 2.1: An EC framework

computing methods by monitoring how natural phenomena act to solve complex problems in different environmental circumstances. The primary goal of EC is to create computational models and efficient algorithms based on natural intelligence discovered in physical, chemical, and biological systems. Among various other optimization techniques, EC methods are proven as a global optimization method for multiple optimization problems. These methods are population-based approaches; hence, they solve any optimization problems by uniformly creating a random initial population and utilizing variation operators (mutation, crossover) for generating the next population or offspring. They also follow the natural selection process for further refinement of the population for the next generation. The basic overview of the EC framework is explained in Figure 2.1.

The EC methods originated in the 1960s when researchers proposed Genetic Algorithm (GA), Genetic Programming (GP), and Evolution Strategies (ES) for finding solutions to global optimization problems [93]. Further, Differential Evolution (DE) and Estimation Distribution of Algorithms (EDA) were other development made in 1990, and all these approaches are known as evolutionary algorithms. 1990 onwards, a few other optimization methods were also designed by simulating natural intelligent phenomena of ants and swarms, etc., called Swarm Intelligence [94]. All these methods are known to the sub-classes of EC [14]. An overview of these categories of EC is shown in Figure 2.2. Because of their derivative-free nature, EC algorithms are generally suited to non-convex and black-box optimization problems where even the mathematical form of the objective function is not available. Thus, they are the center of attention and lead to several creations. A summary of these intelligent algorithms is provided in Table 2.1. These methods are well-suited for multi-objective and many-objective optimization problems also. A few multi-objective evolutionary algorithms are detailed in Table 2.2. Recently, due to the emergence of Big Data, researchers from academia and industry both face 4 “Vs” (Volume, Variety, Velocity, and Value)[7] as critical issues

Table 2.1: Brief summary of intelligent optimization algorithms

Reference	**Algorithm**	**Motivation**	**Year**
J.H. Holland [28]	Genetic Algorithm (GA)	Inspired from Darwin Evolution's theory	1975
Kirkpatrick et.al [29]	Simulated Annealing(SA)	Trajectory based approach inspired from annealing process	1983
Kennedy et al. [30]	Particle Swarm Optimization (PSO)	Smart social behavior of bird flock	1995
Hersovici et al. [31]	Shark Search Algorithm	Feeding mechanism and coordinated movement of fish	1998
Nara et al. [32]	Sheep Flock Heredity Model	Natural evolution of sheep flocks	1999
Passino [33]	Bacterial Foraging Algorithm	Foraging strategy of *E. Coli* bacteria	2002
Li [34]	Artificial Fish Swarm Algorithm (AFSA)	Fish swarm's collective intelligence	2003
Martin et al. [35]	Termite Algorithm	Termite colony	2006
Dorigo [36]	Ant Colony Optimization (ACO)	Behaviour of real ant colony	2006
Karaboga and Basturk [37]	Artificial Bee Colony (ABC)	Honey Bee	2007
Mucherino et al. [38]	Monkey Search (MS)	Monkey's food-seeking behavior while climbing trees	2007
He et al. [39]	Group Search Optimizer (GSO)	Foraging behavior in animals	2009
Yang [40]	Firefly Algorithm	Firefly flashing behavior	2009
Yang and Deb [41]	Cuckoo Search	Obligate brood parasitism of cuckoo	2009
Yang [42]	Bat Algorithm	Bat echolocation behavior	2010
Pan [43]	Fruit Fly Optimization algorithm (FFOA)	Fruit fly's fruit-seeking behavior	2012
Gandomi et al. [44]	Krill Herd (KH)	Krill's herd herding behavior in nature	2012
Kaveh et.al [45]	Dolphin Echolocation	Dolphin echolocation ability	2013
Cuevas et al. [46]	Social Spider Optimization Algorithm	Social spider's cooperative behavior	2014
Uymaz et al. [47]	Artificial Algae Algorithm	Microalgae living behaviors	2015
Mirjalili [48]	Ant Lion Optimizer (ALO)	Ant lions hunting mechanism in nature	2015
Mirjalili et al.[49]	Dragonfly Algorithm	Dragonfly swarming behaviors, both static and dynamic	2016
Abedinia et al. [50]	Shark Smell Optimization(SSO)	Shark's ability to locate prey using its smell sense	2016
Yong et al. [51]	Dolphin Swarm Optimization Algorithm (DSOA)	Mechanism of dolphins to detect, chase, and prey on sardine swarms	2016
Li et al. [52]	Virus Colony Search	Virus infection and diffusion strategies	2016
Mirjalili and Lewis [53]	Whale Optimization Algorithm (WOA)	Humpback whale social behavior	2016
Mirjalili et al. [54]	Multi-verse Optimizer (MVO)	Based on a cosmology theory concept	2016
Askarzadeh [55]	Crow Search Algorithm (CSA)	Crows' intelligent food-hiding behavior	2016
Mirjalili et al. [56]	Salp Swarm Algorithm	Salps' swarming behavior when navigating and foraging in the oceans	2017
Saremi et al. [57]	Grasshopper Optimization Algorithm	Grasshopper swarming behavior	2017
Fausto et al. [58]	Selfish Herd Optimizer (SHO)	Hamilton's selfish theory	2017
Dhiman et al. [59]	Spotted Hyena Optimizer	Spotted hyenas' social behavior	2017
Qi et al. [60]	Butterfly-inspired Algorithm	Butterfly mate-searching mechanism	2017
Jahani et al. [61]	Mouth Brooding Fish Algorithm	Life cycle of mouth brooding fish	2018
Kaur and Arora [62]	Chaotic Whale Optimization	Hybrid approach to improve the efficiency of WOA	2018
Torabi et al.[63]	Improved Raven Roosting Optimization Algorithm	Ravens' social roosting and foraging behavior	2018
Jain et al. [64]	Squirrel Search Algorithm (SSA)	Dynamic foraging behavior	2019
Bharti et al. [65]	Genetic Directed Weighted Complex Network PSO	Dynamic network topology of swarms	2021
Dhiman et al. [66]	Rat Swarm Optimization	Rat chasing and attacking behavior	2021
Salehan et al. [67]	Corona Virus Optimization	Corona virus characteristics and behavior	2021

Table 2.2: Important development of MOEA

Multi-objective Evolutionary Optimizer			
Reference	**Year**	**MOO**	**Description**
Schaffer et al.[68]	1985	VEGA	Vector evaluated genetic algorithm
Fonseca and Fleming[69]	1993	MOGA	Multi objective genetic algorithm
Horn et al.[70]	1994	NPGA	Niched Pareto genetic algorithm
Srinivas and Dev[71]	1994	NSGA	Non dominated sorting genetic algorithm
Zitzler and Theile[72]	1999	SPEA	Strength pareto evolutionary algorithm
Zitzler et al.[73]	2000	SPEA-II	Strength pareto evolutionary algorithm-II
Deb et al.[74]	2000	NSGA-II	Non dominated sorting GA-II
Knowels et al.[75]	2000	PAES	Pareto archived evolutionary strategy
Corne, Knowels and Oates[76]	2000	PESA	Pareto envelope based selection algorithm
Zitzler et al.[73]	2001	SPEA-II	Strength pareto evolutionary algorithm-II
Corne et al.[77]	2001	PESA-II	Pareto envelope based selection-II(region based selection)
Erikson et al.[78]	2001	NPGA-II	Niched pareto GA-II
Okabe T et al.[79]	2002	VEDA	Voronoi-based estimation of distribution algorithm for MOO
Xue and Sanderson[80]	2003	MODE	Pareto based multi-objective differential evolution
Coello Coello[81]	2004	MOPSO	Multi-objective PSO
Pelikan et al.[82]	2005	MOHBOA	Multi objective hierarchical Bayesian optimization algorithm
Coello Coello et al.[83]	2005	MISA	Multi-objective immune system
Zhang and Li[84]	2007	MOEA-D	MOEA based on decomposition
Quingfu et al.[85]	2008	RM-MEDA	Regularity model-based multi-objective estimation of distribution algorithm
Gong et al.[86]	2008	NNIA	Non dominated neighbourhood immune Algorithm
Huang et al.[87]	2009	MOSaDE	Multi-objective self adaptive differential evolution
Chen et.al [88]	2015	NSLS	Nondominated sorting and local search based MOEA
Li et.al [89]	2016	MOSPL	MO Self paced Learning
Lin et.al [90]	2017	NSBLS	Nondominated sorting and bidirectional local search
Liu et.al [91]	2019	MONSGA-II	Multi-oriented optimization heuristic strategy NSGA-II
nouri et.al [92]	2021	MOFOA	Multi-objective forest optimization algorithm

which also raise the complexity of optimization problems. A 5-M concept [95] is used to categorize complex continuous optimization problems into Many-dimensions, Many-optima, Many-changes, Many-costs, and Many-constraints. The link between complex optimization based on 4 “Vs” and 5 “M” is illustrated in Figure 2.3 and gives insight into future directions.

The mutation and crossover operations are based on guided randomness, and the method, as a whole, typically does not require any gradient knowledge about the objective function being optimized. Because of their derivative-free nature, EC algorithms are generally suited to non-convex and black-box optimization problems where even

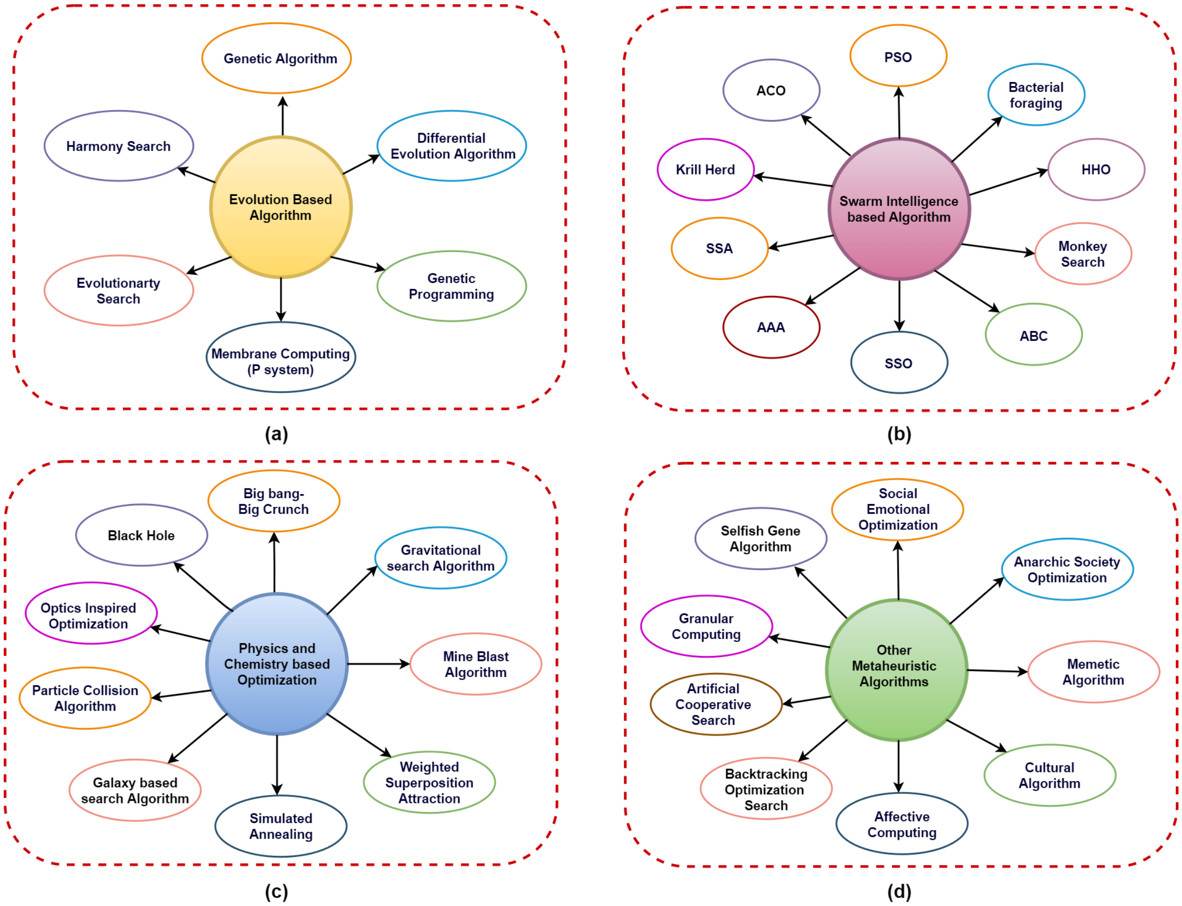

Figure 2.2: Brief overview of EC

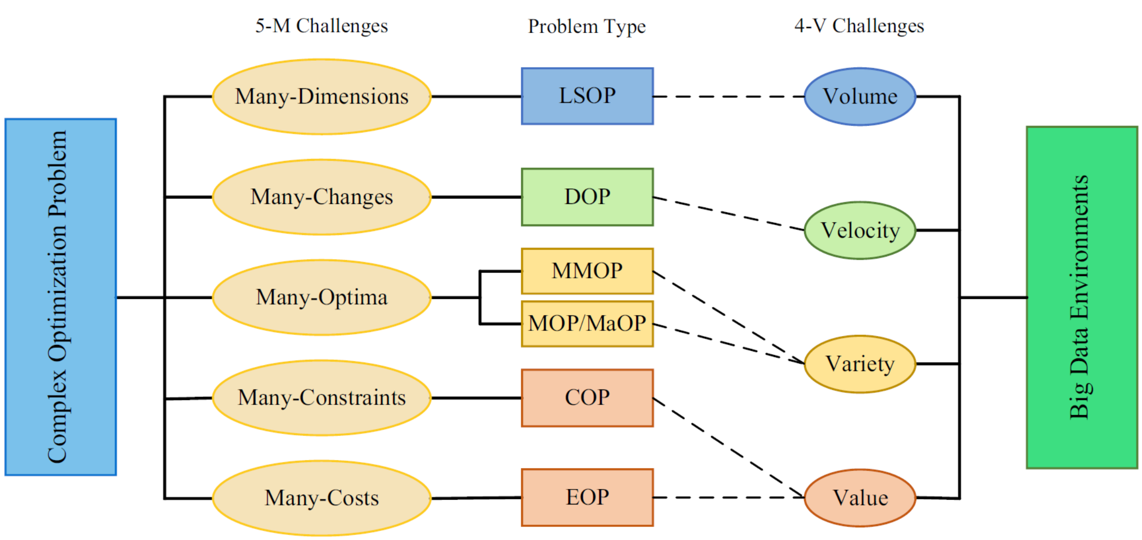

Figure 2.3: The link between complex continuous optimization problems and 5-M and 4-V challenges

the mathematical form of the objective function is not available. The nature of the variation and selection operations distinguishes one EC approach from another. ES, GA, GP, DE, EDA, and so on are the most extensively utilized algorithmic sub-families under EC.

2.1.3 What's new in EC

The recent areas of research within EC are depicted in Figure 2.4. After observing recent research and applications, we have divided the ongoing works into two sub-classes, one of which involves theoretical studies such as convergence analysis, stability analysis, and multidisciplinary studies with EC, and the other belongs to real-world applications of EC embedded ML and deep learning models. Multidisciplinary studies integrated concepts of distinct domains with these EC methods and resulted in new better performing optimizers such as chaos theory from mathematics is integrated with several swarm-based algorithms [96, 97, 98], superposition theorem, big bang theory and black hole from physics give rise to new optimizer [99, 100]. Recent trends in this domain show that research communities are attracted to multidisciplinary studies and their integration with models to improve the quality of solutions and try to reduce exhaustive computation. Moreover, several research demonstrates that researchers are now solving the real-world and challenging applications by treating them as an optimization problem that can be a single-objective, many-objective, dynamic objective, hierarchical in nature, and many more [101, 102, 103].

Furthermore, the latest work demonstrates that research communities are continuously striving to develop promising and effective optimizers inspired by natural phenomena, animal behavior, and physical science [104, 105]. Apart from these, the new popular research area emerges as an introduction of chaos theory in EC. Chaos theory has random, dynamic, non-repetitive, and ergodic properties. Due to its dynamic property, it ensures the different solutions given by algorithms even on a complex multi-modal

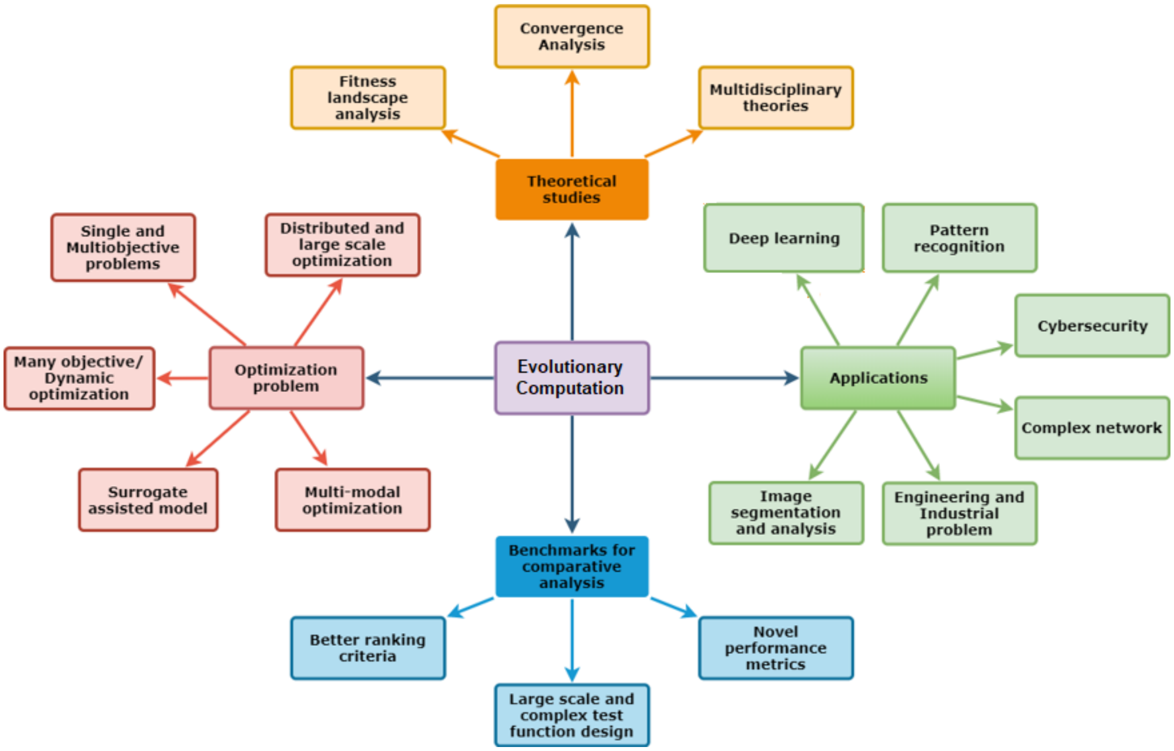

Figure 2.4: Graphical representation of the areas within EC and some of their recent applications

landscape. Since most of the EC belongs to stochastic algorithms in which randomness is achieved by using probabilistic theory such as gaussian distribution or uniform distribution. Nowadays, to accelerate convergence and enhance the diversity of the EC, chaos theory is used. Due to its ergodicity, it performs searches at a higher speed. This novel approach has been widely utilized in various EC methods, and their applications [106]. Another challenging area is convergence analysis which still needs to be explored. Only a few works are there that tried to solve this by using the Markov chain model [107]. Hybrid model of existing EC to enhance the performance of optimizer and ML embedded with EC [13] are few recent developments which are used either for feature selection or model parameter optimization and solving several real-world applications [108, 109]. Self-adaptive EC [110], cellular automata-based EC [111, 112], surrogate assisted EC [113] and new innovation in multi-objective optimization and many-objective optimization are the new attraction for solving versatile applications [114, 115]. There is continuous innovation in technology resulting in more complex problems and the need for better optimizers. It also leads to the design of new complex benchmarks, such as high-scale benchmarks for validating optimizers [116]. In the last few months, it has been observed that deep learning, which is highly popular in almost every domain of application, also integrates with EC, and here it comes a new complex, interesting and challenging area to be explored more. Besides, advanced technology such as Quantum computing, Parallel computing, and GPU recently provided a new edge to EC and ML. These techniques accelerate the development of methods suitable for them, and it is still in the exploration phase and needs to do more in this direction.

2.1.4 Deep Learning: An Overview

Deep learning is a prominent technique in ML and AI, and it has evolved into a core learning method of the revolutionary industry, Industry 4.0. Deep learning is emanated from the neural network and due to its exceptional learning capabilities from data make it worthy and popular among industries and research communities. However, neural networks were widely acceptable models in ML and AI in late 1980s. Following that, various innovative models like multi-layer perceptron trained through backpropagation, radial basis function (RBF) networks, etc., were developed. However, after a certain time, researchers were losing their interest in the neural network. Again, in 2006 Hinton et al. [117] proposed a new generation neural network which is basically a rebirth of neural networks in the form of deep learning, and now it is accepted as a more successful technique in every field. The rise of deep learning has a long history which was further divided into i) First generation had a time period (1958- 1969) that started with a single-layer perceptron neural network developed by Rosenblatt [118]. But once the observation made by Minsky [119] in 1969 was presented that single-layer perceptron was incapable of solving problems that are linearly inseparable, resulting in the research on the neural network being halted for around 20 years. ii) Second generation lay in

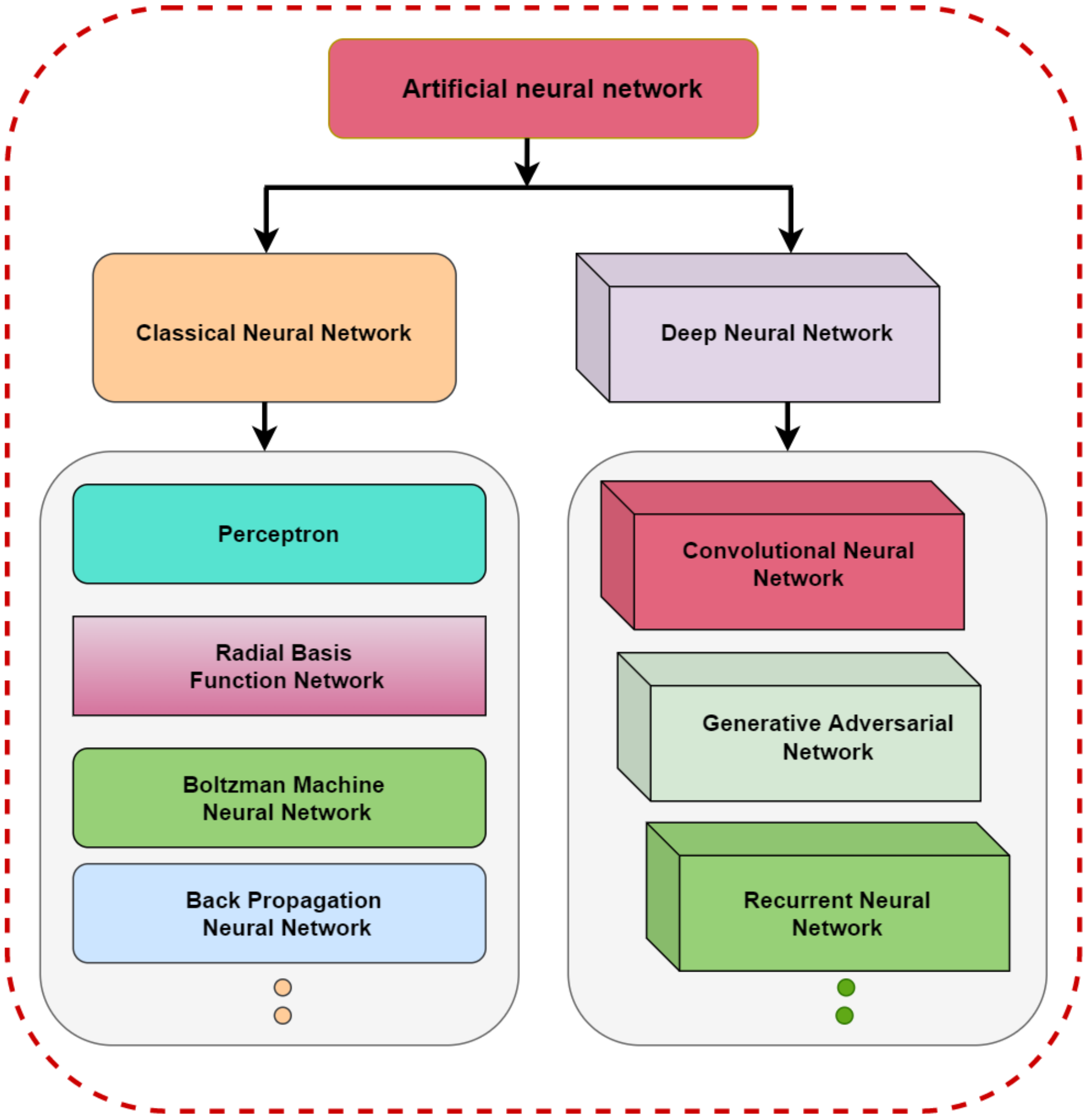

Figure 2.5: Broad categorization of neural network

1986-1998 where Hinton et al. [120] proposed multi-layer perceptron with backpropagation, sigmoid function for nonlinear embedding, which is capable of solving nonlinear classification problems. Further, Hornik [121] came up with a universal approximation theorem, and after that, LeCun et al. [122] proposed a Convolutional Neural Networks (CNN), but it takes three days for training, and in 1991 gradient vanishing problem was reported. Besides, unclear theoretical and mathematical foundations and trial-error approaches weaken the usability of this model while other statistical learning methods with strong mathematical foundations gain popularity. Finally, the third generation neural network models (2006 to present) change the perception of researchers when a graphical model of the brain is explored and presented in the form of a deep learning model [123]. After that, several inventions like Auto Encoders (AE), deep belief network, AlexNet, VGGNet, and many more are presented. Even to overcome the problem of vanishing gradient, solutions like pre-training and then tuning were given. Authors in [124] also proposed a ReLU activation function, which is capable of suppressing the gradient vanishing problem. Due to technological innovations like GPU, deep learning has become more popular in different domains. Deep learning refers to a collection of methods and models that includes, but is not limited to, LSTM, CNN, GAN, AE, and many others. The broad classification of an artificial neural network is presented in Figure 2.5.

2.1.5 Evolutionary Computation with Deep Learning

The performance of deep learning models is influenced by other factors like topology structure, hyper-parameters, learning rules, data quality, and optimum weights. Optimizing the deep learning architecture with enhanced performance for specific applications like image classification is a tedious task. However, to make it an efficient model, a vast knowledge of deep learning and image processing is needed, which is usually not possible for researchers of other interdisciplinary areas. Therefore, several trial and error experiments are conducted for parameter tuning as well as to get the best architecture for specific applications, and similar architecture can not be suitable for different applications. Hence an intelligent mechanism is needed to automate the whole process for better adaptability. That intelligent mechanism must be flexible, efficient, derivative-free, and easy to understand such that it can be extended to the different architectures of deep learning models. Thus, the most suitable mechanisms for such tasks are EC techniques. It is currently the most active research domain, attracting significant interest from both academia and industry. An overview of the application

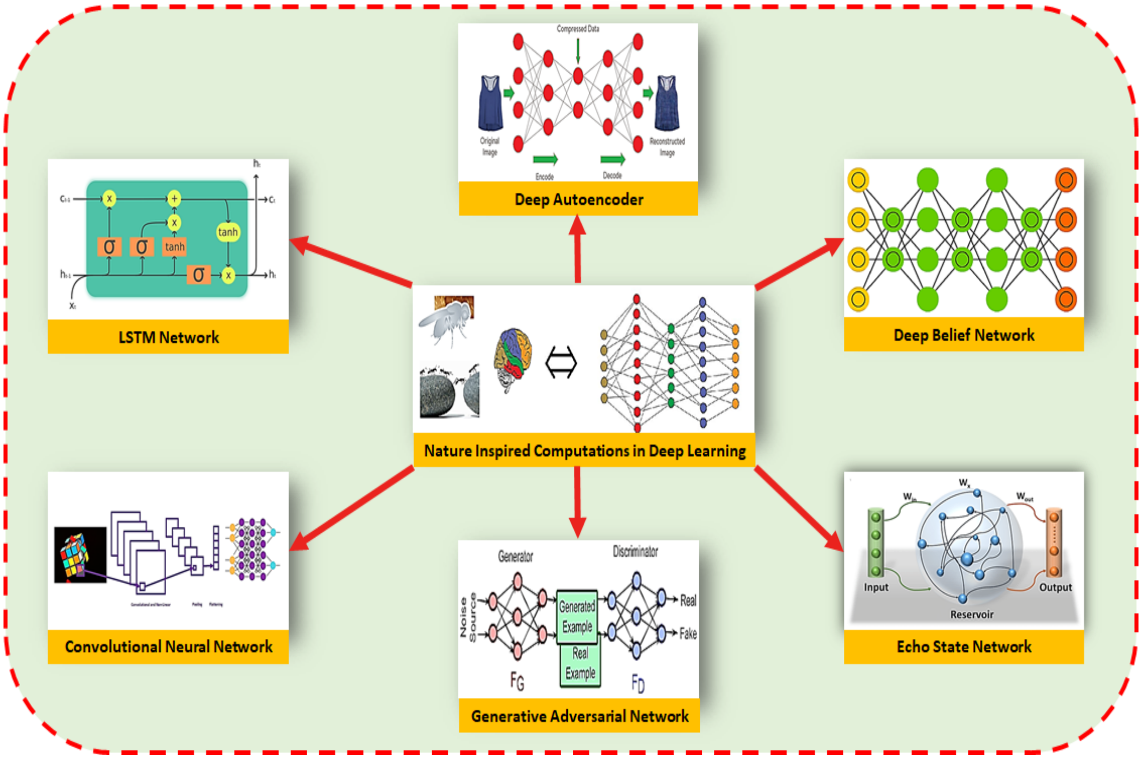

Figure 2.6: Applications of EC in deep learning.

of EC in deep learning is presented in Figure 2.6. The AI concepts of neuro-evolution, introduced by "Stanley and Miikkulainen" [125] utilizing deep learning, have recently attracted considerable attention. It refers to the use of simulated evolution to build artificial neural networks, including their learning mechanism, optimal weights, and network structure. Hence, the optimization of deep learning models using EC techniques is a topic of interest and debate in computer science and other fields. In deep learning, we focused primarily on Evolutionary GAN, and the related work of GAN with EC is presented in the next section.

2.2 Related Works

2.2.1 EC and Data Security

The evolution of Industry 4.0 gives a competitive edge to industrial applications, and almost every industry has benefited, including healthcare, agriculture, digital imaging, and so on. This breakthrough facilitates data transfer over a public network while also introducing new challenges and opportunities in a variety of fields. Among these challenges, data security and integrity have recently risen to the top of almost every industry's priority list, including finance, medicine, and the smart digital world. Furthermore, with the widespread development and advancement of IoTs [126], sensors and fast connectivity across the telecommunications system provide us with comfort and convenience, especially in the medical sector, for patient reports and sensitive data transmission. Thus, in order to facilitate high-quality healthcare services at a low cost, IoT-assisted mobile cloud-based e-health services are making great strides by leveraging new technologies such as big-data, medi-cloud, blockchain, and IoT in healthcare [127]. To accomplish these goals, IoTs play a dominant role in the development of smart HealthTech [128]; however, such systems suffer from the additional risk of data theft and security breach. The cyber attacks are mainly targeting the essential national infrastructure like banks, hospitals, and the electric power grid, which use and rely on SCADA and industrial control systems to handle their operations [5]. On June 14, 2017, a ransomware attack gained access to the sensitive data of 266,123 patients at the Pacific Alliance Medical Center in Los Angeles, and other similar incidents have been reported [129, 130]. Therefore, the frequent cases of data manipulation attacks such as health information manipulation, tampering, and information theft make data privacy and security one of the most difficult aspects of smart healthcare.

Hence, researchers addressed this problem and provided various techniques, such as watermarking, encryption, compression, and steganography [131, 132, 133]. In addition, few research works have also been identified for healthcare data security within the IoT network. One of these works [134] has introduced a security microvisor middleware (SμV) for the representative IoT device, which enables customized security operations and memory isolation using software virtualization and low overhead assembly code verification in terms of memory and battery life. Similarly, Manogaran et al. in [135] proposed a secure industrial IoT-based Meta cloud redirection(MC-R) framework to collect the data from different sensors. These medical sensors are embedded in the human body for the collection of clinical measurements such as heart rate, body temperature, respiratory rate, blood pressure, and blood sugar of patients. If these measures exceed their normal value, a warning message comprising these clinical measures shall be sent to doctors through the wireless sensor network. They used a key management security mechanism to ensure the transmission of large data.

Moreover, very recently in [136], authors have presented a novel blockchain, i.e., PUFchain, which integrates hardware security with primitive physical unclonable functions, hashing module, and blockchain for robust as well as enhanced data security, and device security. They also introduced a new consensus algorithm with PUFchain that can easily be integrated into a resource-constrained IoT setting in order to address IoT energy needs, scalability, and latency. In another work, a reversible interpolation-dependent watermarking technique based on GA and PSO is developed and applied to medical as well as standard data [137]. Similarly, grasshopper optimization with GA is used to create an optimal key that has been applied to both data sanitization, and restoration processes [138]. After analyzing its convergence and comparing it with other methods for medical data, the authors found this to be more effective than others.

Encryption is another possible solution, and because these techniques are based on standard cryptography, they are widely used to ensure the secure storage of sensitive data, such as medical records and reports. Traditional encryption techniques include “Data Encryption Standard (DES)” [139], “triple DES (3DES)” [140], “Advanced Encryption Standard (AES)” [141], etc. However, these techniques are best suited for text data, whereas medical images are mostly used for analysis. Several initiatives have been launched in this direction. The authors [127] proposed a grasshopper-PSO hybrid algorithm that was used to improve the security of the encryption and decryption process for medical data by selecting the optimal key. In another work [142], proposed a novel Elliptic Galois Cryptography (EGC) based on cryptography and steganography for secure data transmission over an IoT network. The properties of the elliptic curve theory were used to generate the key, and an EGC protocol with adaptive firefly optimization was proposed for hidden and secure data in IoTs. The mean square error, carrier capacity, and peak signal-to-noise ratio were all used.

Recent work presented in [143] also presented a resource optimization model for clinical data transmission with low processing time and energy consumption, as well as a biometric-based security model that extracts heartbeats from ECG signals. In this method, a unique biometric key was generated using block encoding to generate binary bits, and binary features were extracted from the encoded bits of the heartbeat. These sequences were concatenated and evaluated as a biometrically generated key, which was initially 128-bit and sufficient to encrypt the original medical information using logical operations, enabling efficient and secure transmission between patients and remote doctors.

Authors [144, 145] have proposed techniques for preserving a single medical image based on serial mode. Further, another work [146] used high-speed scrambling and pixel adaptive diffusion to encrypt the medical image. To improve the diffusion effect, a bi-directional adaptive technique [144] changed all pixel values in a single round of diffusion. In [145], medical images were secured using hierarchical diffusion in a non-sequential manner after evaluating the bit distribution of medical images. They have shown experimentally that only two rounds of encryption were capable of delivering satisfactory encryption. Low dimensional chaotic maps are easy to implement, but they have limitations like narrow intervals and few parameters, which are addressed in another work [147], for medical data encryption with the help of 2D chaotic maps. A recent study [148], presented a faster and more practical cryptosystem for image encryption using parallel computing and chaotic encryption, where they used permutation and substitution architecture of chaotic encryption.

2.2.2 EC for Feature Subset Selection

ML techniques are widely used for discovering meaningful patterns and classifying real-world data. These datasets may be large and complex, so feature selection is the primary strategy for reducing the dimension of the data, with the general goal of reducing the amount of redundant and disruptive features in a dataset for fast and efficient data analysis without sacrificing significant predictive model performance. Due to exponentially high search space, feature selection is a complex optimization problem. It is practically impossible to evaluate all of the feature subsets manually. Various search techniques, such as random, greedy, complete, and heuristic, have been utilized for feature subset selection [149, 150]. Based on the evaluation criteria, feature selection is generally classified into three categories: (1) filter-based approaches, (2) wrapper-based approaches, and (3) embedded approaches [151]. However, there are no detailed guidelines on the benefits and drawbacks of alternative approaches. An overview of feature selection using EC is shown in Figure 2.7.

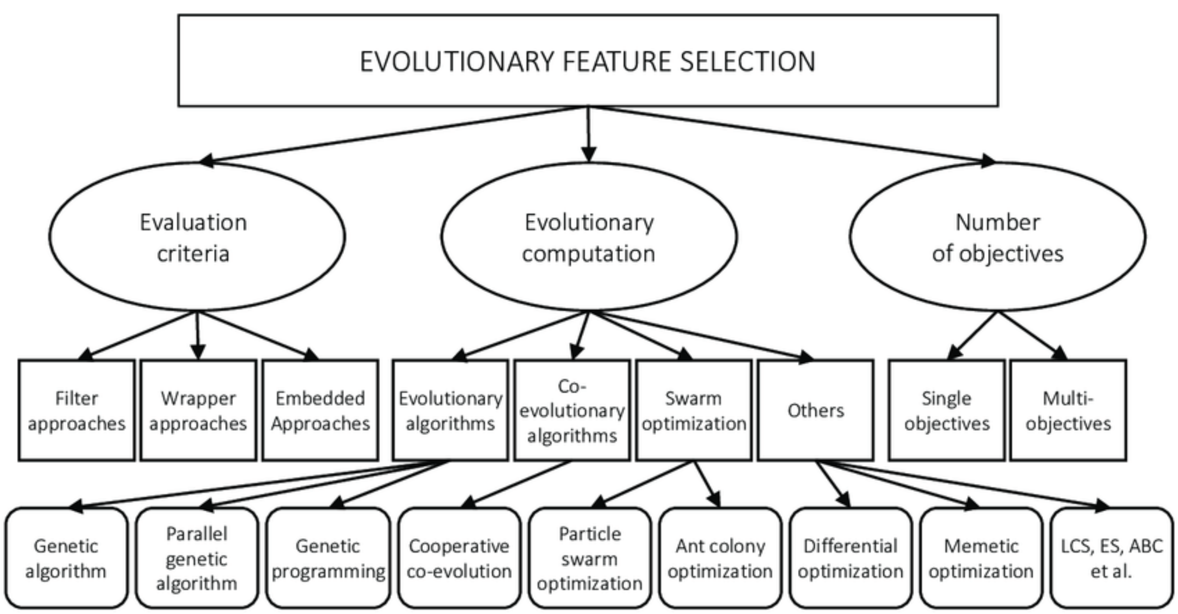

Figure 2.7: Overview of feature selection approaches

Large search space and possible feature interaction make classical greedy and other Non-EC approaches, like filter-based methods, susceptible to local optima. These methods are computationally faster than EC approaches, but in the case of large-scale data computing, the probabilistic value of quality and providing rank to each feature is infeasible. They are good selectors for small-scale datasets; however, these days, due to industrialization and advanced technology, voluminous datasets are generated and give rise to the large-scale optimization problem.

On the other hand, EC is widely used for this due to its inherent capability of conducting global search and proven effectiveness for feature selection problems. They are usually known as the wrapper-based approach, where these methods utilize a learning model as an evaluator within a loop. They evaluate the entire feature subset while accounting for possible feature interaction and perform better than other approaches. The main issue, however, is their frequent premature convergence, resulting in an inadequate contribution to data mining. Even the majority of existing optimizers are not adaptive. In [152], the process of feature selection in various domains using EC is briefly studied, along with the challenges involved. Generally, most wrapper methods suffer from local optima stagnation and high computational cost. To overcome these issues, an efficient global search technique must be designed.

Furthermore, scalability is another challenge due to the emergence of big data [7]. In early 1989, a dataset with more than 20 features and the selection of features from this data was referred to as large-scale feature selection [153]. However, the number of features in several areas, such as gene analysis, has grown from thousands to millions in recent years, increasing the computation cost and necessitating an advanced search mechanism, both of which have their own limitations. As a result, simply having a lot of computing power isn't solve the problem. Designing novel methods and searching mechanisms for handling this type of complex optimization problem is an emerging area of research. Although several contributions and applications have been developed from diverse domains for feature selection using EC approaches. EC methods solved feature selection problems either as a single objective problem or a multi-objective problem.

GA and PSO are widely used in the application domain for high-dimensional feature selection tasks. Using NSL-KDD [154], the GA is combined with bagging and a partial decision tree to select optimal features and classification for an intrusion detection system. Similarly, GA with Bayesian Network achieved 98.26% accuracy with reduced 16 features [155]. Another work [156] also used GA for feature selection and parameter optimization of SVM in the intrusion detection system.

Recently [157] also shows the credibility of GA in epilepsy seizure detection using SVM and ANN. In another study [158], authors introduced a new initialization and update mechanism in PSO for feature selection with SVM. Further variable length PSO [159] is also proposed for high-dimensional feature selections. In another recent work [160], a synergistic method for estimating state-of-health (battery life) achieves 95% accuracy using the GA and support vector regression.

Overall, EC has a wide range of applications for feature selection tasks. Table 2.3 provides a brief description of contributions made by researchers in wrapper-based feature selection from the last few years.

2.2.3 EC with GAN

GAN is a deep learning model that has proven to be a powerful image generation tool in recent years. It is made up of two parts: the generator (G) and the discriminator (D). The G learns how to generate data that looks similar to genuine training data in order to fool the D, whereas the D learns how to distinguish between data from a real training set and data generated by the G.

Table 2.3: Related works for wrapper-based method for feature selection.

Ref.	Year	Datasets	EC Method	Classifier	Performance Metrics	Conclusion
[161]	2014	SRBCT, MLL Leukemia, and Acute Leukemia	An improved regularized linear discriminant (Improved RLDA)	Four classifiers called J4.8, NB, KNN (K = 1), and SVM	Accuracy and No. of features.	Improved RLDA produces the best results when employing the SVM classifier, with an accuracy of 100% for all three datasets utilizing 10% of the genes.
[162]	2014	11 Tumors, 14 Tumors, 9 Tumors, Brain Tumor 1, Brain Tumor 2, Leukemia 2, Lung Cancer, SRBCT, and Prostate Tumor	PSO	DT (C4.5)	Accuracy and No. of features.	In six datasets, the proposed approach beats all existing methods and has an accuracy rate of more than 90%.
[163]	2015	Colon, Prostate	A modified version of the Inertial Geometric Particle Swarm Optimization (MIGPSO)	SVM, Self-Organizing Map, back propagation neural network, NB, DT, Artificial Immune Recognition System, and PSO with C4.5 Decision Tree	Accuracy, No. of features, time.	When compared to existing approaches, the presented approach enhances classification accuracy by roughly 4%. It achieves an accuracy of 95.8% with 6 genes in the Colon dataset and 98.04% with 84 genes in the Prostate dataset.

Continued on next page

Table 2.3 – *Related works for wrapper-based method for feature selection*

Ref.	Year	Datasets	EC Method	Classifier	Performance Metrics	Conclusion
[164]	2015	Prostate Cancer, Leukemia, and Colon Tumor	Modified PSO (MPSO)	SVM, KNN, and NB	Precision, Recall, F-Score, AUC, Accuracy, and No. of features.	Using an SVM classifier on a Prostate cancer dataset with 20 genes, MPSO achieves an accuracy of more than 99% and an F-score of 90%. With 10 genes and an SVM classifier, MPSO achieves an accuracy of 99.01% for the remaining datasets.
[165]	2016	Four microarray datasets were taken from the UCI repository, which are Wisconsin Breast Cancer (WBC), Colon Cancer, Lymphoma, and Leukemia	HPSO-LS, which is a particle swarm optimization with a local search strategy	KNN	Accuracy and running time in seconds.	With four genes, HPSO-LS achieves an accuracy of 98.08% for the WBC dataset.
[166]	2017	Nine datasets that include one microarray dataset of SRBCT	Artificial fish swarm (AFSO)	SVM	Accuracy and No. of features.	The proposed methodology achieves an accuracy of 82% while employing 215 features for the SRBCT dataset.

Continued on next page

Table 2.3 – *Related works for wrapper-based method for feature selection*

Ref.	**Year**	**Datasets**	**EC Method**	**Classifier**	**Performance Metrics**	**Conclusion**
[167]	2017	Seven microarray datasets - Cancer (Lung, Ovarian, Prostate, Breast, Colon), Leukemia (ALL/AML), and DLBCL.	Nero-fuzzy with firefly algorithm for tuning parameters	Rule-set generation as classifiers	Accuracy and No. of features	For the Lung cancer and Ovarian cancer datasets, the proposed approach has an accuracy of 93.42% with 4 genes and 96.13% with 12 genes.
[168]	2018	CNS, Breast cancer, Colon tumor, Prostate Cancer, Diffuse Large B-Cell Lymphoma (DLBCL), Leukemia, Lung cancer-Michigan, Lung cancer, Ontario,	Cuttlefish Algorithm	KNN, DT, Hidden Markov models, and SVM	Accuracy, No. of features, and time in seconds.	The method described reduces the number of features by up to 90%. It achieves greater than 92% accuracy in 7 of 8 datasets while taking less time to compute than other approaches. The accuracy for the Prostate dataset approaches 100%.
[169]	2018	Colon-Tumor, Leukemia (ALL/AML), Lung Cancer	GBC	MBP-CGP architecture is used with one and two hidden layers	Accuracy and No. of features	The algorithm selects features ranging from 47-51% for all datasets, and its accuracy ranges between 88.75 and 100% for all datasets.

Continued on next page

Table 2.3 – *Related works for wrapper-based method for feature selection*

Ref.	Year	Datasets	EC Method	Classifier	Performance Metrics	Conclusion
[170]	2018	ALL-AML, MLL, CNS, SRBCT, Breast and Lymphoma	iBPSO	SVM, NB, NN (K = 1)	Accuracy and No. of features	The presented technique using NB yields the lowest average number of genes, which is 120.83. In comparison to SVM, and KNN at K = 1, it gives an average of 138.83 and 133.66 genes, respectively. The method's average highest classification accuracy with KNN, NB, and SVM is 90.58, 93.98, and 93.02, respectively.
[171]	2018	Breast Cancer Wisconsin dataset	Binary teaching learning-based optimization algorithm	NB, KNN, DT, SVM, and Discriminant Analysis	Precision, Recall, Sensitivity, F-Measure, Accuracy, Receiver Operating Characteristic (ROC) curve, and No. of features	Using NB, SVM, and Discriminant Analysis classifiers, the suggested technique achieves an accuracy of more than 98% and an F-measure of almost 97%.
[172]	2019	Ten microarray datasets, including three high-dimensional datasets	GOFS	SVM and KNN	Accuracy and No. of features.	GOFS achieves 100% accuracy in the Lymphography dataset.

Continued on next page

Table 2.3 – *Related works for wrapper-based method for feature selection*

Ref.	Year	Datasets	EC Method	Classifier	Performance Metrics	Conclusion
[173]	2019	GDS531, and GDS2643	ABC-SDS	SVM	Accuracy, sensitivity, specificity, and F score	For the GDS531 dataset, ABC-SDS outperforms ABC alone by 9.53%, 9.01%, 9.01%, and 9.48% in terms of accuracy, sensitivity, specificity, and F score, respectively. For GDS2643, ABC-SDS outperforms ABC alone by 10.75%, 10.04, 10.04, and 11.88% in terms of accuracy, sensitivity, specificity, and F score, respectively.
[174]	2019	Eight microarray datasets	Binary bat algorithm	ELM with a new fitness function	Accuracy, precision, recall, specificity, F score, and No. of features.	The presented feature selection technique achieves 100% accuracy in 6 of 8 datasets, with a feature selection percentage of 25.80% on average.
[175]	2019	SRBCT, Lung, Colon, Leukemia1, and Leukemia2	Firefly algorithm	SVM	Accuracy and No. of features	FF-SVM achieves 100% classification accuracy across the Leukemia1 dataset with three genes and the Lung dataset with two genes. It has an accuracy of 95.2% and 83.3% over SRBCT and Leukemia2, respectively, when employing five genes.

Continued on next page

Table 2.3 – *Related works for wrapper-based method for feature selection*

Ref.	Year	Datasets	EC Method	Classifier	Performance Metrics	Conclusion
[176]	2019	Seven microarray datasets (SRBCT, Colon, Prostate, MLL, DLBCL, AMLGSE2191, etc.)	RMA	Three well-known classifiers (SVM, Multi-Layer perceptron (MLP), and KNN).	Accuracy and No. of features.	The results reveal that RMA gives 100% accuracy in all circumstances while using a minimal no. of genes.
[177]	2020	One microarray dataset (Breast cancer) from the UCI repository	BWOA	Logistic Regression, C4.5, Naïve Bayes (NB)	Accuracy and No. of features	The proposed methodology achieves the highest mean accuracy with percentages of 97% (Logistic Regression), 98% (C4.5), and 97% (NB) for the Breast cancer dataset in a 50-50 training validation test.
[178]	2020	Eight microarray datasets, including Breast cancer, Breast EW	Monarch butterfly optimization algorithm (MBO)	KNN	Accuracy, recall, precision, specificity, F-score, and No. of features	MBO surpasses four meta-heuristic algorithms: WOASAT, ALO, GA, and PSO, which choose less than 40% of features from all datasets while achieving 100% accuracy, recall, precision, specificity, and F-score in 5 of 8 datasets.
[179]	2022	SRBCT, Prostate tumor, Brain tumor	ChOA	KNN, NB, NN	Accuracy, number of features,	ChOA achieves maximum accuracy 98.70% with NB on SRBCT, 84.3% with KNN on Prostate tumor and 87.7% with NB on a Brain tumor.

Following training, the G has a high likelihood of producing data that is useful in a number of real-world applications. It started a new revolution in deep learning and computer vision. This revolution has resulted in some notable advances in research. Hence this section investigates an EC footprint for GAN's generator training improvement. Generally, the training of the G can be expressed in the form of an optimization problem. Although the fitness function for the G cannot be directly created, its local state may be assessed as a function of D performance. So, the overall training process is basically a min-max optimization problem. Recently, the potential of intelligent computation techniques has also been experimented on this framework.

E-GAN [180] was the first proposal in this direction where they design training of GAN with different evolutionary operators, and a number of generators (population) are trained. They tried to give a solution for stable training that could alleviate the problem of mode collapse. Further, authors in [181] have introduced a novel model incorporating neuroevolution and co-evolution in the GAN training method. They employed the loss function (fitness function) for the discriminator and the Frchet Inception Distance (FID) for the generator. Further, they experimented on the MNIST dataset and showed that for the generator, the FID score is a good evaluation metric. However, it is crucial to measure competing metrics for discriminators. It is further extended by the author in [182], presenting an enhanced model CO-EGAN based on neuroevolution and coevolution techniques by incorporating the adversarial feature of GAN components in the training process to construct coevolutionary methods.

Further, EvoGAN [183] is used for facial image generation, CG-GAN (Composite generating GAN)[184] for facial composite generation, and finally, multi-objective EGAN [185] have been proposed. MO-EGAN further improved in [186] by improving the training of a single objective generator by using a multi-criteria training process. Recently, the authors in [187] presented GANs-PSO that tackle typical GAN problems such as mode collapse. In this work, GANs-PSO used the PSO algorithm with GAN to overcome the mode collapse problem during training for self-collision avoidance of arm robots. While PSO is also applied in [188], where authors designed a GAN by utilizing the concept of neuroevolution for biomedical image chest X-ray of COVID-19 generation. The whole training process, along with the architecture search, was achieved using PSO. FID score is used as a fitness value, and a progressive growth approach is utilized to design GAN. The suggested approach achieves a superior FID score for image generation as compared to the baseline approach. Another work [189] utilized fractional Harris Hawk optimization to optimize GAN training for Osteosarcoma detection at an early stage to enhance the survival rate. PSO is also utilized in [190] to optimize the GAN for face generation, and based on quality and diversity evaluation metrics, the position of particles is updated.

2.3 Summary

Nature is incredibly interesting and yet to be interpreted fully. Undiscovered paradigms underneath natural science will certainly proceed to foster new developments in EC, with improved performance and computing effectiveness. This chapter offers a theoretical foundation of underlying models, recent trends, and related work. EC involves setting parameters on its own, and the best systematic way of achieving optimal parameter values for EC remains an open research issue. Additionally, slow convergence, local optima stagnation, poor diversity, non-adaptiveness, and scalability are major issues with EC, limiting their applicability to large-scale optimization problems and other modern applications. Because of these concerns, the potential of EC is underutilized, necessitating efficient search mechanisms and approaches to address their bottlenecks. Furthermore, future applications will require the use of advanced computing techniques and resources with these methods to reduce the cost of the evaluation.

The addition of EC in deep learning (advanced ML model), therefore, constitutes extra parameter values. However, the parameter values of the deep learning models can be decreased as some of the parameters can be automatically determined by the EC algorithms. This would have a multiplier effect on deep learning in a situation where the EC parameter values are not significant to deliver high performance. This could explain why the model obtained by a search became stuck in local minima. Furthermore, the main concerns of generative models are mode collapse, vanishing gradient, and training instability. The same concerns exist with Cyclic-GAN, and addressing these concerns is still in the course of the investigation.