IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 2 February 2023, accepted 19 March 2023, date of publication 29 March 2023, date of current version 5 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3263198

SURVEY

Portfolio Optimization Problem: A Taxonomic Review of Solution Methodologies

ZI XUAN LOKE[1], SAY LENG GOH[1], GRAHAM KENDALL[2], SALWANI ABDULLAH[3], AND NASSER R. SABAR[4]

[1]Optimisation and Visual Analytics Research Group, Faculty of Computing and Informatics, Universiti Malaysia Sabah Labuan International Campus, Labuan 87000, Malaysia
[2]School of Computer Science, University of Nottingham Malaysia, Semenyih, Selangor 43500, Malaysia
[3]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor 43600, Malaysia
[4]Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia

Corresponding author: Say Leng Goh (slgoh@ums.edu.my)

This work was supported by Universiti Malaysia Sabah under Grant SBK0463-2021.

ABSTRACT This survey paper provides an overview of current developments for the Portfolio Optimisation Problem (POP) based on articles published from 2018 to 2022. It reviews the latest solution methodologies utilised in addressing POPs in terms of mechanisms and performance. The methodologies are categorised as Metaheuristic, Mathematical Optimisation, Hybrid Approaches, Matheuristic and Machine Learning. The datasets (benchmark, real-world, and hypothetical) utilised in portfolio optimisation research are provided. The state-of-the-art methodologies for benchmark datasets are presented accordingly. Population-based metaheuristics are the most preferred techniques among researchers in addressing the POP. Hybrid approaches is an emerging trend (2018 onwards). The OR-Library is the most widely used benchmark dataset for researchers to compare their methodologies in addressing POP. The research challenges and opportunities are discussed. The summarisation of the published papers in this survey provides an insight to researchers in identifying emerging trends and gaps in this research area.

INDEX TERMS Hybrid approaches, machine learning, mathematical optimisation, matheuristic, metaheuristic.

I. INTRODUCTION

Investors usually hold multiple assets in an investment portfolio [1]. A portfolio is a composition of financial assets (stocks, funds, bonds, or commodities). Diversifying the portfolio is common among investors to maximize the expected return while minimizing the risk [2]. The Portfolio Optimisation Problem (POP) is one type of combinatorial optimisation problems (COPs) [3], [4], [5]. It involves assigning resources to a limited number of assets, which seeks an equilibrium between risk and return. In 1952, Harry Markowitz proposed a Mean-Variance (MV) model in addressing the POP. The model aims to either maximise the return of a portfolio while defining the lowest acceptable risk or to minimise the risk of a portfolio while defining the lowest acceptable return, as a linear constraint [6]. As returns are associated with different risk levels and the risk tolerance for each investor is variable, there is no universal portfolio that is acceptable for all investors [7]. The MV model provides a fundamental understanding on how to optimize a portfolio. In a real-world setting (complicated and volatile financial market), cardinality constraint, bound constraint, transaction costs etc. are considered in generating a practical portfolio.

An optimal portfolio is important to determine a combination of financial assets that suits a specific investor. This helps investors make the right decision by selecting and allocating the optimal proportion of assets from the financial market to increase their profit. In 1991, Markowitz was awarded the Nobel Prize (economics) for the recognition of his pioneering contributions to the financial theory. Due to his contribution, the POP started gaining attention from researchers in a variety of domains, including mathematicians and computer scientists [8], [9], [10].

There are several survey papers related to POP in the scientific literature. Aouni et al. surveyed different Goal Programming (GP) models such as polynomial GP, lexicographic

The associate editor coordinating the review of this manuscript and approving it for publication was Jerry Chun-Wei Lin.

GP, weighted GP and fuzzy GP that were used to solve the POP [11]. Masmoudi and Abdelaziz reviewed objective deterministic and stochastic programming models for POP, where various key concepts of POP were summarised and the mathematical models for solving the models were presented [12]. Kalayci et al. reviewed deterministic models and implementations of MV models in a detailed manner, which included variants of the model and the real-world constraints. The survey categorised the solution methodologies into approximate and exact methods and examined the existing algorithms based on a variety of performance indicators [13].

As these survey papers were published many years ago, it is timely to review the latest work for the POP and update the taxonomy of POP. The contributions of this survey paper are;

- The definition, variants, and real-world constraints of the POP are provided. The recent solution methodologies are systematically categorised (Metaheuristic, Mathematical Optimisation, Matheuristic, Hybrid Approaches, and Machine Learning) and analysed (mechanisms and performance). The benchmark, real-world, and hypothetical POP datasets are identified and presented as well as the relevant state-of-the-art methodologies.
- Solution methodologies are presented on a timeline (2018 to 2022) to identify emerging trends. Furthermore, the methodologies are tabulated by benchmark dataset to discover popular datasets that are currently in use.
- Research challenges and opportunities are discussed.

The rest of the paper is structured as follows: Section II gives the scope and methodology of this survey. Section III describes the POP. Section IV reviews the solution methodologies to address the POP. Section V presents the research challenges for the problem. Section VI describes the dataset available for the problem. Section VII suggests the future directions and the conclusion is given in Section VIII.

II. METHODOLOGY AND SCOPE OF SURVEY

This survey aims to document the evolution of the POP and its solution methodologies in recent years. Around 80 methodological articles on POP published between 2018 to 2022 were collected from various digital libraries (Google Scholar and Scopus) using the keywords such as "Portfolio Optimisation", "Portfolio Selection", "Financial Market", and "Optimisation Problem". These articles were distributed in subject areas such as computer science, decision science, mathematics, and engineering. 42 articles published in top-tier journals were selected for analysis. Figure 1 displays the distribution of articles for each journal while Figure 2 exhibits the spread of articles per year.

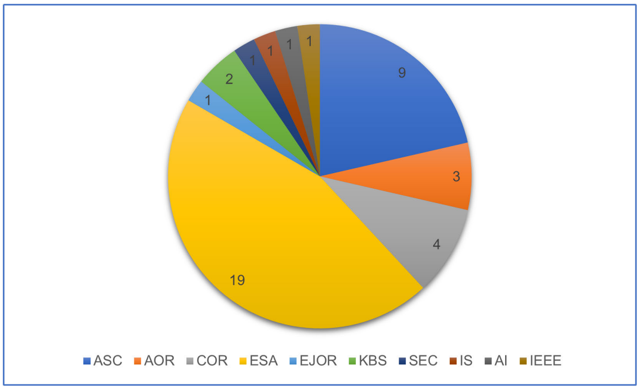


FIGURE 1. Distribution of articles for each Journal. Applied Soft Computing (ASC), Annals of Operations Research (AOR), Computers and Operations Research (COR), Expert Systems with Applications (ESA), European Journal of Operational Research (EJOR), Knowledge-Based Systems (KBS), Swarm and Evolutionary Computation (SEC), Information Sciences (IS), Journal of Ambient Intelligence and Humanized Computing (AI), and IEEE Access (IEEE).

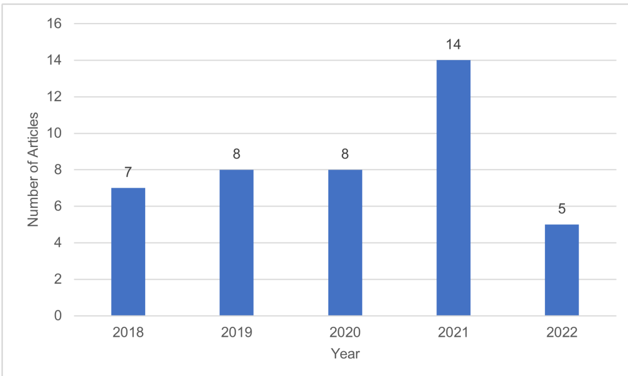


FIGURE 2. Distribution of articles per year.

III. PORTFOLIO OPTIMIZATION PROBLEM

A. PROBLEM DEFINITION

Portfolio optimisation is a decision-making process where a fixed sum of money is distributed among a variety of financial assets, and the weight allocation of the assets is constantly changed to maximize the return and minimize the risk [6]. The main goal of the portfolio optimisation is to determine which assets and the proportion of the selected asset to invest in, considering the available amount of money.

B. PROBLEM VARIATIONS

Two variants of POP are found in the scientific literature; single-objective and multi-objective.

1) SINGLE-OBJECTIVE POP

The classical POP (Markowitz mean-variance model) consists of only a single objective, where investors choose to either minimize portfolio risk or maximize portfolio return. In minimizing the portfolio's risk, the risk (measured by portfolio variance) is minimized (Equation 1), subject to a minimum return specified by an investor (Equation 2), and two other constraints; weights of assets sum up to one (Equation 3) and range between zero and one (inclusive) (Equation 4). N represents the total number of assets available, μ_i is the expected return of an asset i, R^* indicates the minimum desired return, w denotes the respective weights of the portfolio's assets, and σ_{ij} is the covariance between asset i and asset j.

$$\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} w_i \times w_j \times \sigma_{ij} \tag{1}$$

subject to

$$\Sigma_{i=1}^{N} w_i \times \mu_i = R^* \tag{2}$$

$$\Sigma_{i=1}^{N} w_i = 1 \tag{3}$$

$$0 \leq w_i \leq 1; \quad i = 1, \ldots, N \tag{4}$$

2) MULTI-OBJECTIVE POP

There are two approaches in handling multi-objective POP: weighted sum and pareto-based. In a weighted sum approach, a set of objectives is integrated into a single objective by assigning a corresponding weight to prioritize one over another [14]. Multi-objective POP can be formulated as follows using a weighted sum approach (Equations 5 to 7) [1]. The two conflicting objectives (minimizing risk, maximizing return) are weighted using a parameter, λ. When λ = 0, the model aims to maximize the return of the portfolio (ignoring the portfolio's risk). When λ = 1, the model seeks to minimize the risk of the portfolio (ignoring the portfolio's return). The λ parameter achieves a trade-off between risk and return (0 < λ < 1).

$$min\ \lambda[\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} w_i \times w_j \times \sigma_{ij}] - (1-\lambda)[\Sigma_{i=1}^{N} w_i \times \mu_i] \tag{5}$$

$$\text{subject to } \Sigma_{i=1}^{N} w_i = 1 \tag{6}$$

$$0 \leq w_i \leq 1; \quad i = 1, \ldots, N \tag{7}$$

In a Pareto-based approach, a multi-objective POP is formulated as in Equations 8-10. The two conflicting objectives (minimizing risk, maximizing return) are independently assessed to achieve a Pareto front (a set of optimal solutions)

$$min\ [\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} w_i \times w_j \times \sigma_{ij}] and\ max\ [\Sigma_{i=1}^{N} w_i \times \mu_i] \tag{8}$$

$$\text{subject to } \Sigma_{i=1}^{N} w_i = 1 \tag{9}$$

$$0 \leq w_i \leq 1; \quad i = 1, \ldots, N \tag{10}$$

C. REAL-WORLD CONSTRAINTS

Some practical constraints are included in the portfolio model (problem formulation) to cover real-world factors and limitations. The constraints are;

- *Round-lot Constraint*: Certain assets can only be purchased in multiples of a certain quantity (lots).
- *Pre-assignment Constraint/ Contingent Constraint*: An investor's preference for a particular collection of assets.
- *Cardinality Constraint*: Limit the quantity of an asset in a portfolio.
- *Inequality Constraint*: This constraint is the same as the cardinality constraint. However, it is specified in terms of lower and upper bounds.
- *Floor and Ceiling Constraint*: Determine the investment proportion (upper and lower bounds) for each asset, which reflect the preference of an investor for specific assets, a group of assets or economic sectors.
- *Transaction Costs Constraint*: A fee paid by the investors when buying or selling stocks.
- *Turnover Constraint*: Measure the liquidity of an asset using the ratio between the market's average stock and the outstanding shares (tradable stock) of the asset.
- *Bankruptcy Constraint*: A situation where an investor's wealth falls below some predetermined level and investment must be terminated to avoid potential losses and credit risks.

IV. CATEGORISATION OF SOLUTION METHODOLOGIES

In this section, we present the methodologies utilized by researchers to address the POP. Table 1 shows the solution methodologies sorted in a chronological order (according to year). The most recent solution methodologies are; Non-Linear Activated Beetle Antennae Search [15], Multi-Objective Genetic Algorithm [16], Improved Genetic Algorithm [17], Quadratic Programming [18], and a combination of Variable Neighborhood Search and Monte Carlo Simulation [19].

Figure 3 shows the categorisation of solution methodologies for the POP. These methodologies can be categorised as Metaheuristic, Mathematical Optimisation, Hybrid Approaches, Machine Learning, and Matheuristic.

Table 2 exhibits the breakdown of the solution methodologies by category and year.

There are 20 metaheuristics (2 single solution-based and 18 population-based), 5 mathematical optimisations, 12 hybrid approaches, 4 machine learning techniques and 1 matheuristic. Metaheuristics (especially the population-based variant) are the most popular approach in addressing the POP. This is followed by hybrid approaches, mathematical optimisation, machine learning and matheuristic. Matheuristic is the least used approach among the solution methodologies. The use of hybrid approach appears to be growing in popularity from 2018 to 2021.

A. METAHEURISTICS

1) SINGLE SOLUTION-BASED ALGORITHMS (SS-MH)

a: BEETLE ANTENNAE SEARCH

Khan et al. applied a Beetle Antennae Search (BAS) algorithm in addressing a constrained POP (NASDAQ stock exchange data) [35]. A constrained POP model (Markowitz's mean-variance model) was formulated. The model was improved by adding cardinality and transaction cost constraints. The algorithm was inspired by the food searching behaviour of a beetle which uses a pair of antennae to measure the intensity of smell in various directions when searching for food. The beetle moves towards a direction with stronger smell intensity. The proposed methodology was comparable to, and faster than, Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), and Pattern Search (PS).

Khan et al. proposed a Non-Linear Activated Beetle Antennae Search (NABAS) algorithm to address the non-convex tax-aware POP (real-world data from NASDAQ stock

TABLE 1. Solution methodologies for the POP (According to year).

Year	Methodology	Reference
2018	Multi-objective Evolutionary Algorithm (MOEA)	[20]
	Multi-objective Evolutionary Algorithm (MOEA)	[21]
	Mixed Integer Linear Programming (MILP)	[22]
	Polynomial Goal Programming (PGP)	[23]
	Fuzzy Goal Programming (FGP)	[24]
	Firefly Algorithm (FA), Simulated Annealing Algorithm (SA)	[25]
2019	Non-dominated Sorting Genetic Algorithm II (NSGA-II)	[26]
	Adaptive Ranking Multi-Objective Particle Swarm Optimisation (AR-MOPSO)	[27]
	Improved Multi-objective Particle Swarm Optimisation (IMPSO)	[28]
	Information Based Evolutionary Algorithm (Info EA)	[29]
	Polynomial Goal Programming (PGP)	[30]
	ARMA-GARCH Econometric Model, Support Vector Regression (SVR)	[31]
	Electromagnetism-like (EM) Algorithm, Genetic Algorithm GA), Genetic Network Programming (GNP), Particle Swarm Optimisation (PSO),Simulated Annealing (SA) (All the algorithms are hybridised with diversification mechanism)	[32]
	Support Vector Machine (SVM)	[33]
2020	Non-Dominated Sorting Genetic Algorithm II (NSGA-II)	[34]
	Beetle Antennae Search (BAS)	[35]
	Non-dominated Sorting Genetic Algorithm (NSGA), Neural Network Decision Maker 2 (NNDM-2)	[36]
	Two-stage Clustering (TSC), Radial Basis Function (RBF) Neural Network (NN), Genetic Algorithm (GA)	[37]
	Continuous Ant Colony Optimisation (CACO), Genetic Algorithm (GA), Artificial Bee Colony (ABC)	[38]
	Variable Neighborhood Search (VNS), Quadratic Programming (QP)	[39]
	Long Short-Term Memory (LSTM)	[40]
	Stacked Deep Dynamic Recurrent Reinforcement Learning (SDDRRL)	[41]
2021	Modified Squirrel Search Algorithm (MSSA)	[42]
	Genetic Algorithm (GA)	[43]
	Genetic Algorithm (GA)	[44]
	Genetic Algorithm (GA)	[45]
	Genetic Algorithm (GA)	[46]
	Modified Version of Non-dominated Sorting Genetic Algorithm II (MNSGA-II)	[47]
	Improved Quantum-Behaved Particle Swarm Optimisation Algorithm (LQPSO)	[48]
	Artificial Bee Colony (ABC)	[49]
	Power Activation Feed-Forward Neuronet (PFN), Two-stage Beetle Antenna Search (BAS)	[50]
	Valence Aware Dictionary and sEntiment Reasoner (VADER), Google's Bidirectional Transformer for Financial Data (FinBERT), Genetic Algorithm (GA)	[51]
	Genetic Algorithm (GA), Non-Linear Neural Network (NNN)	[52]
	eXtreme Gradient Boosting (XGBoost), Improved Firefly Algorithm (IFA), Monte Carlo Method	[53]
	Non-dominated Sorting Genetic Algorithm II (NSGA-II), Genetic Algorithm (GA)	[54]
	Random forest (RF), Support Vector Regression (SVR), Long Short Term Memory (LSTM), Neural Network (NN), Deep Multilayer Perceptron (DMLP), Convolutional Neural Network (CNN)	[55]
2022	Non-Linear Activated Beetle Antennae Search (NABAS)	[15]
	Multi-Objective Genetic Algorithm (MGO)	[16]
	Improved Genetic Algorithm (IGA)	[17]
	Quadratic Programming (QP)	[18]
	Variable Neighborhood Search(VNS), Monte Carlo Simulation (MCS)	[19]

market) [15]. A gradient estimate measure was used as an update criteria in a BAS algorithm. As it decreased exponentially, an activation threshold was proposed to overcome premature convergence. The threshold allowed searching particles of the algorithm to fully explore the search space at a given gradient estimate measure, preventing the algorithm from getting stuck in a local minima. A GlobalSearch() function in MATLAB was used to evaluate the fitness of the algorithm. NABAS performed better than BAS, and it was comparable to Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO).

2) POPULATION-BASED METAHEURISTIC ALGORITHMS (P-MH)

a: GENETIC ALGORITHM

A Genetic Algorithm (GA) was proposed by Gupta et al. to address multi-period fuzzy POP (data from three

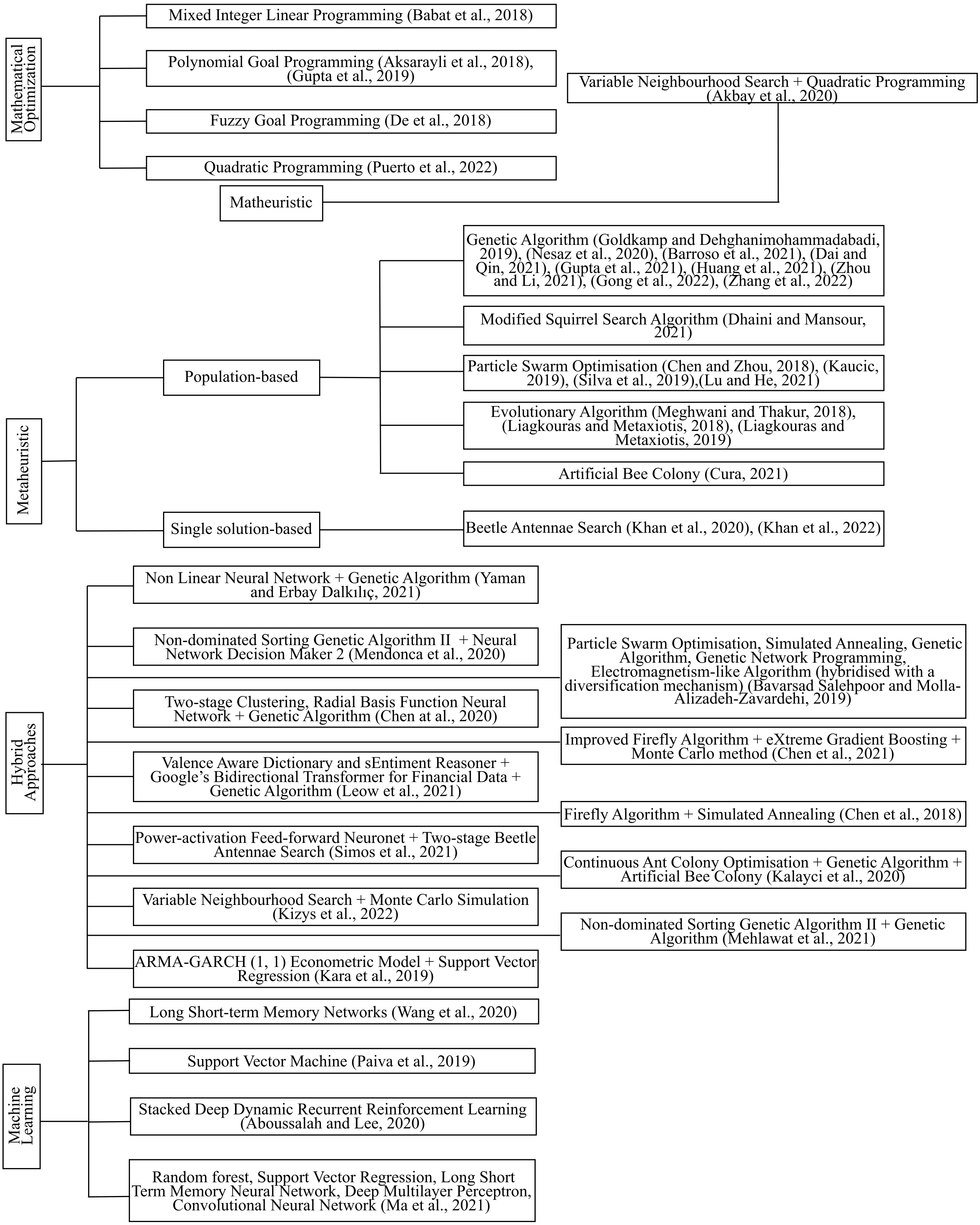


FIGURE 3. Categorisation of solution methodologies for the POP.

TABLE 2. A breakdown of the Solution Methodologies by Category and Year. SS-MH: Single Solution-based Metaheuristic, P-MH: Population-based Metaheuristic, MO: Mathematical Optimisation, HA: Hybrid Approach, ML: Machine Learning, M: Matheuristic.

Year	SS-MH	P-MH	MO	HA	M	ML	Total
2018	0	3	3	1	0	0	7
2019	0	4	1	2	0	1	8
2020	1	1	0	3	1	2	8
2021	0	8	0	5	0	1	14
2022	1	2	1	1	0	0	5
	2	18	5	12	1	4	42

real-world stock markets, National Stock Exchange (NSE) for the first case study, S&P 500 and NASDAQ-100 indexes for the second case study) [45]. Two portfolio models were presented; multi-period multi-objective mean-mean absolute semi-deviation (MASD)-skewness and multi-period multi-objective mean-Conditional Value at Risk (CvaR)-skewness portfolios. A coherent fuzzy number (an extension of fuzzy number) was used to capture the returns of asset and the perception (optimistic, neutral, pessimistic) of investors in the stock market. There were three objectives for the model; maximizing mean, minimizing MASD or CvaR, and maximizing skewness. The investor first chose one objective to be the model's objective function and treated the other objectives as the constraints of the model. Asset allocations corresponds to solutions in the algorithm. Roulette wheel selection was used to choose parents (to create offspring for the next generation). The offspring were iteratively improved by simulated binary crossover and non-uniform mutation operators. An elitism operator was used to preserve the best-fit solutions to ensure there was an improvement in each iteration. A constraint handling method was used to ensure the feasibility of the offspring. The algorithm stopped searching when an investor was satisfied with the solution. The proposed portfolios were comparable to a naive portfolio. The authors concluded that examining the current stock market prospects (perception of investors) accurately was an important consideration in addressing POP.

Dai and Qin applied a Genetic Algorithm (GA) to address the multi-period uncertain POP (data from Shanghai Stock Exchange (SSE)) [43]. A multi-period mean-Value-at-Risk (VaR) portfolio model was proposed, where returns of the stocks were regarded as uncertain variables. The goal of the model was to maximize the return while treating the risk as a constraint of the model. An initial population was initialized by creating a random set of feasible chromosomes. Roulette wheel selection was used to generate a parent population. Crossover and mutation operators were used to produce offspring. A feasible repair mechanism was performed to ensure the feasibility of the chromosomes. The GA could solve the proposed model effectively as the algorithm converged to the optimal value within a small number of generations.

Gong et al. applied an Improved Genetic Algorithm (IGA) [56] to address two multi-period fuzzy POP (stocks from CSI100 index (first dataset) and SSE180 index (second dataset)) [17]. Two multi-period fuzzy mean-variance portfolio models were proposed. The goal of the first model was to maximize return and treat risk as one of the constraints of the model. The goal of the second model was to minimize risk and treat return as one of the constraints of the model. Coherent fuzzy numbers were used to capture the uncertainty of asset returns and the different attitudes (optimistic, pessimistic, neutral) of investors towards the stock markets. After parameter initialisation, an initial population of individuals were randomly created. Laplace crossover and power mutation operators were utilised in crossover and mutation operations respectively. Roulette wheel selection was utilised. A constraint-handling operation was performed to ensure the feasibility of individuals. The process was iterated until a stopping criterion was satisfied. The proposed models outperformed traditional static models in terms of objective functions and evaluation criterias.

Zhang et al. applied a multi-objective genetic algorithm (MOGA) to address POP with sell orders (using data from two real-world stock markets, New York Stock Exchange (NYSE) for the first case study and Hong Kong Stock Exchange (HKSE) for the second case study) [16]. A portfolio model with sell orders was integrated with an automatic trading system to generate an investment strategy. The model's objective was to maximize portfolio return and minimize its risk. Weighted fuzzy frequency (WFF) was used to determine the fuzzy return of an investment strategy as an LR-power fuzzy number, based on historical data. The credibility expected value and semi-variance were used to measure return and risk of the investment strategy. An encoding method was used to satisfy the constraints of the model. All candidate solutions were categorized into different dominance classes. Solutions in the same dominance class were evaluated according to their isolation degree. The higher the dominance class and isolation degree, the fitter a solution. The risk-return pair of all the efficient solutions form the Pareto optimal solution (efficient frontier) of the investment strategy. Finally, the fuzzy Value-at-Risk (VaR) ratio for all the efficient solutions were calculated, and the efficient solution with the highest fuzzy VaR was selected as the optimal investment strategy. Results showed that MOGA outperformed genetic algorithm (GA) and Particle Swarm Optimisation (PSO) in terms of least running time.

Huang et al. employed a Genetic Algorithm (GA) to tackle an uncertain POP with background risk (BR) (hypothetical dataset [57]) [44]. BR referred to the non-financial market risk caused by different factors, including changes in health expenditure, real estate investment, and labour income. An asset that is exposed to BR was called a background asset. An uncertain risk index portfolio optimisation model with background risk (RIMWBR) was proposed. The return of assets and background assets were considered as unknown variables and predicted by experts. The model's objective was to maximize return, where risk minimization was treated as one of the constraints of the model. There were four phases in GA: initialization, selection, crossover, and mutation. After initializing initial feasible chromosomes, the chromosomes were selected by roulette wheel selection. Then, the chromosomes were iteratively improved by crossover and mutation operations. Phase two and phase three were repeated for 20 times and the best chromosome (highest return) was returned as the solution of the portfolio. The chromosomes were always checked to ensure its feasibility. The algorithm was tested with different parameters. The GA was efficient and stable in addressing the proposed model. The authors claimed that it was more effective than the traditional mathematical optimisation method.

Zhou and Li presented a Genetic Algorithm (GA) to tackle a multi-period mean-semi-entropy portfolio optimisation model (hypothetical dataset) [46]. The returns of assets were described using fuzzy variables. Fuzzy techniques were beneficial in financial modelling when the future conditions of a financial market could not be accurately predicted by historical records. Transaction costs and bankruptcy control were input as the real-life constraints. An initial population of feasible chromosomes was randomly generated in the initialization process. Roulette wheel selection was employed so that a fitter chromosome has a higher chance of being selected to produce offspring. A crossover operator altered the programming representation of a chromosome while a mutation operator preserved the genetic diversity of chromosomes from one generation to another. Results showed that the proposed approach is practical for multi-period portfolio optimisation.

Nesaz et al. proposed a Non-dominated Sorting Genetic Algorithm II (NSGA-II) to address a multi-period portfolio model with Lower Partial Moment (LPM) (data from New York Stock Exchange (NYSE)) [34]. A new portfolio-driven method (PDM) for calculating the LPM was presented and compared with the stock-driven method (SDM). In NSGA-II, crossover and mutation procedures were utilized for generation of new offspring. A Taguchi experimental design approach was used to obtain optimal parameter values (crossover rate, mutation rate, mutation rate percentage, initial population and number of iteration). PDM was better than SDM in computing LPM in terms of efficiency and processing time in reaching an optimal solution.

Barroso et al. proposed a modified version of Non-dominated Sorting Genetic Algorithm II (NSGA-II) to address a multi-objective (risk, return) portfolio optimisation model (data from Brazilian Stock Market) [47]. Among the real-life constraints considered were cardinality, floor and ceiling. Covariance and Conditional Value at Risk (CVaR) were used as measures of risk. An integration between Technical Analysis (AT) indicators and Portfolio optimisation (OT) was proposed. Two types of integration were considered: OTAT and ATOT. In OTAT, an optimal investment portfolio was generated at the beginning of each month and AT indicators were used to perform transactions. In ATOT, optimisation was performed monthly on selected assets (filtered by the indicators). In NSGA-II, individuals were divided into boundaries of dominance. Individuals were selected based on the concept of dominance and agglomeration distance. A tournament selection was performed, where two individuals were drawn, and the individual with best dominance frontier was chosen. If the two individuals belong to the same dominance frontier, agglomeration distance was used to evaluate the density of solutions. A higher agglomeraton distance represented a more diverse solution (efficient frontier). The integration of OT and AT could provide investors with optimal strategies (higher return for a given amount of risk, considering real-life constraints).

Goldkamp and Dehghanimohammadabadi proposed a multi-objective genetic algorithm (NSGA-II) to address the S&P 500 dataset [26]. An intelligent system that could suggest profitable pair combinations of assets was presented. The pair formation problem was formulated as a Mixed Integer Programming (MIP) model. A selection process was included at the end of each iteration to choose the fittest individuals in creating new candidates for the next generation. Crossover and mutation processes were carried out to improve the candidate solutions. The algorithm generated a Pareto front of solutions, where a decision maker can choose a solution from a range of trading opportunities.

b: SQUIRREL SEARCH ALGORITHM

Dhaini and Mansour proposed a modified Squirrel Search Algorithm (mSSA) to address the unconstrained and constrained POP, using stocks from benchmark datasets, OR-Library and Johannesburg Stock Exchange (JSE) [42]. Cardinality and boundary constraints were considered in the constrained POP. The objectives of the portfolio models were to maximize return and minimize risk. A flying squirrel was used to represent a candidate solution. An initial population of candidate solutions was randomly initialized. A normalization method (for unconstrained POP) and arrangement algorithm (for constrained POP) were used to ensure the feasibility of the solutions. The candidate solutions were evaluated and sorted in an ascending order according to their fitness values. The fittest candidate solution was declared on hickory tree (optimal food source). The following three candidate solutions were declared on acorn trees. The remaining candidate solutions were declared on normal trees with no food source. New candidate solutions were generated as squirrels gliding from one food source to another; from normal trees to acorn trees, from normal trees to hickory trees and from acorn trees to hickory trees. A mathematical equation was utilised in each case to update the new solutions. A gliding constant and a random gliding distance were used in the equation. The gliding constant was used to balance between exploration and exploitation in the equation. The foraging behaviour of the squirrels was affected by weather changes (squirrels were less active in winter compared to autumn [58]). A season monitoring condition was checked to determine if winter season had ended. A random relocation of squirrel was only carried out if the winter season had ended. The relocation was only performed on the squirrel who could not move towards the optimal food source. The algorithm was repeated until a maximum number of iterations. The fittest candidate solution located on the hickory tree was the returned solution. The proposed algorithm was suitable in addressing the unconstrained POP. The mSSA obtained good results for small sized datasets.

c: PARTICLE SWARM OPTIMISATION

Lu and He presented an improved quantum-behaved particle swarm optimisation algorithm (LQPSO) to address a fuzzy POP (data from Hong Kong Stock Market) [48]. The goal was to minimize portfolio risk, while treating return as a constraint. The LQPSO algorithm was an improved version of QPSO [59]. Among the improvements were; an intensive short-range exploration and an occasional long-range search were developed based on rank walk of Levy flight mechanism, a non-linear structure of contraction-expansion coefficient was introduced to regulate the speed of particles and a diversity function was proposed to improve the exploration cability. The proposed algorithm outperformed Genetic Algorithm (GA), Simulated Annealing (SA), Immune Algorithm (IA), Differential Evolution (DE), PSO and QPSO in terms of smaller optimum (mean) and standard deviation.

Kaucic proposed an improved multi-objective particle swarm optimisation (IMOPSO) algorithm to tackle risk parity based cardinality constrained POP (RP-CCPOP) (stock data from Dow Jones Industrial Average, Fama and French 49 Industry, and NASDAQ 100) [28]. The proposed portfolio model was an extension of the Markowitz mean-variance model. Budget constraint, cardinality constraint, buy-in threshold (floor and ceiling) constraint, and risk parity conditions were handled at the same time. Three hybrid techniques; a repair mechanism and different constrained domination principles [60], self-adaptive tolerance constrained domination method [61], and self-adaptive penalty constrained domination method [62] were used to handle the constraints. A swap mutation operator was used to enhance the exploration capability of the algorithm. The integration of repair mechanism with self-adaptive tolerance constrained domination method offered the best compromise between quality of solutions and number of optimal risk-return profiles discovered on approximated Pareto front.

Silva et al. proposed an adaptive ranking multi-objective particle swarm optimisation (ARMOPSO) to address different variants of mean-variance portfolio models (benchmark dataset from OR-Library) [27]. An initial swarm of particles was randomly generated according to the objectives and constraints of the variants (Mean-variance POP, Cardinality Constrained POP, Inequality Cardinality Constrained POP, Inequality Cardinality Pre-assignment Constrained POP, and Cardinality Round-lot Pre-assignment Constrained POP). An adaptive ranking procedure was developed combining three mechanisms: non-dominated sorting (NS), crowding distance (CD), and cost-benefit (CB), to update the non-dominated solutions and determine the best global position in every iteration. A mutation process was conducted to diversify the swarm in each iteration. Different feasibility mechanisms were implemented to repair infeasible solutions during execution of the algorithm. ARMOPSO could obtain highly competitive results in all variants and in most of the adopted performance metrics (median, mean percentage error, minimum, maximum, diversity metric, spacing, hypervolume, generational distance, variance of returned error, mean return error, and error ratio).

Chen and Zhou applied a Multi-objective Particle Swarm Optimisation (MOPSO) to tackle robust multi-objective portfolio model with higher moments (real-world data from the Chinese Stock Market) [63]. A global repository was used by each particle in MOPSO to store its flight experiences (speed and position) after each flight cycle. The repository was updated based on a geographically-based system (defined by the objective function values of each particle). Each particle used the repository to select a leader to lead its search. The proposed algorithm was superior to Non-dominated Sorting Genetic Algorithm II (NSGA-II) and Strength Pareto Evolutionary Algorithm 2 (SPEA2) in terms of spacing metric and average running time.

d: EVOLUTIONARY ALGORITHM

An evolutionary algorithm (EA) includes two populations of individuals as shown in Algorithm 1. The first population is called an archive. It retains the best solutions found during a search. The second population stores a population of offspring which participates in reproduction. Archive A^0 is set to an empty set and population B^0 is set to a random sample of the solution space (line 2). At each generation, individuals from A^t and B^t were evaluated (line 4). The archive A is updated (line 5). Sample and vary operators are used to specify a particular selection and a reproduction scheme (line 6). Finally, the best solution is returned by the algorithm [64].

Algorithm 1 Evolutionary Algorithm [64]

1 $t = 0$
2 $(A^0, B^0) = \text{initialize}()$
3 **while** *termination* $=$ *false* **do**
4   evaluate(A^t, B^t)
5     $A^{t+1} = \text{update}(A^t, B^t)$
6     $B^{t+1} = \text{vary}(\text{sample}(A^t))$
7     $t = t + 1$
8 **end**
9 return best(A^t, B^t)

Meghwani and Thakur modified three Multi-objective Evolutionary Algorithms (MOEA); Non-dominated Sorting Genetic Algorithm II (NSGA-II), Decomposition based MOEA (MOEA/D), and Global Weighting Achievement Scalarizing Function Genetic Algorithm (GWASFGA) to tackle three multi-objective (risk, return, transaction cost) portfolio optimisation models with different risk measure (variance, Value-at-Risk (VaR) and Conditional Value-at-Risk (CvaR) (FF38 and FF48 datasets from Fama and French Data Library) [21]. Quantity, pre-assignment, cardinality, and transaction cost were included in the portfolio model as real-life constraints. A repair algorithm was introduced to handle all equality constraints. Roulettle wheel selection was used to choose an asset among candidate assets at each iteration. The probability of selection was based on the past returns of assets. The NSGA-II and GWASFGA were comparable. Both outperformed MOEA/D algorithm in terms of hypervolume performance metric.

Liagkouras and Metaxiotis proposed an information based evolutionary algorithm (Info Based EA) to tackle a fuzzy portfolio model (real-world FTSE-100 dataset) [29]. An initial solution was randomly generated. Genetic operators (selection, crossover, and mutation) were used to improve the initial solution. A vector was proposed to track changes made to the solution during the mutation process. Good solutions were reproduced using the information stored in the vector. The proposed algorithm was better than Non-dominated Sorting Genetic Algorithm II (NSGA-II) in terms of three performance metrics; Hypervolume indicator, Inverted Generational Distance, and Epsilon indicator.

Liagkouras and Metaxiotis proposed a multi-objective evolutionary algorithm (MOEA) to address a cardinality constrained POP (OR-Library dataset) [20]. An encoding scheme was used to represent a portfolio. A real-valued vector stored information about the chosen assets and the budget allocated. A binary vector indicated the existence of a certain asset in the portfolio. The size of both vectors were equivalent to the upper limit of the cardinality constraint. A two-phase mutation operator (polynomial mutation (PLM) operators were applied on the real-valued vector while a bit-flip (BF) mutation operator was applied on the binary vector) and a two-phase recombination operator (simulated binary crossover (SBX) operators were applied on the real-valued vector while a single point crossover operator was applied on the binary vector) were introduced. The proposed approach outperformed Non-dominated Sorting Genetic Algorithm II (NSGA-II) and Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D).

e: ARTIFICIAL BEE COLONY

Cura proposed an Artificial Bee Colony (ABC) algorithm to address seven publicly available benchmark problems (five from OR-Library and one from XU030 and XU100 indices each) [49]. A cardinality constrained POP model was presented. In the ABC, there were three types of bees: employed, onlooker, and scout. The employed bees moved between two randomly chosen food sources based on the amount of nectar. The onlooker bee randomly chose a food source based on a probability (amount of nectar) associated with the food sources. The scout bees collected information from all the solutions. A new solution was generated by the scout bees using the information if a solution could not be improved within a predetermined number of iterations. The proposed approach was compared with seven other approaches: Genetic Algorithm (GA), Tabu Search (TS), and Simulated Annealing (SA) (GTS) [1], an Improved Particle Swarm Optimisation (IPSO) algorithm [65], a hybrid of population-based incremental learning and differential evolution algorithms (PILD) [66], a GRASP-based algorithm [67], an ABC algorithm with feasibility enforcement and infeasibility toleration procedures (ABCFEIT) [68], a hybrid of ABC, Ant Colony Optimisation (ACO), and GA (AAG) [38], and a hybrid of Variable Neighbourhood Search Algorithm and Quadratic Programming (VNSQP) [39]. The proposed ABC outperformed all the seven approaches in terms of efficiency.

B. MATHEMATICAL OPTIMISATION

1) INTEGER PROGRAMMING

Babat et al. proposed two algorithms (Algorithm A and Algorithm B) that exploited a Mixed-Integer Linear Programming (MILP) formulation to solve a Value-at-Risk (VaR) portfolio model (US stock market data obtained from Kenneth R.French's Website) [22]. These two algorithms were developed to generate near-optimal solutions. Algorithm A was executed iteratively to find a good feasible solution. Algorithm B was used to increase the optimality of the feasible solution. The two proposed algorithm achieved better performance than Gurobi (MILP solver) in addressing the VaR portfolio problem (hundreds of assets and thousands of samples).

2) POLYNOMIAL GOAL PROGRAMMING

Gupta et al. proposed a polynomial goal programming approach to solve fuzzy mean-variance-skewness-entropy (MVSE) portfolio model (real-world National Stock Exchange dataset) [30]. Among the real-world constraints considered were cardinality, floor, ceiling and contingent. The preferences of investors and market conditions were represented by floor, ceiling and contingent constraints. Four objectives (mean, variance, skewness, entropy) were considered in the model. Intuitionistic fuzzy set (IFS) was applied to construct the extreme (minimum and maximum) values of these objectives. Decision makers could obtain different results based on their preferences. The proposed approach was better than other approaches [69], [70].

Aksaraylı et al. applied a polynomial goal programming (PGP) approach to solve a portfolio model with higher moments (data from Istanbul Stock Exchange (ISE) and 12 industry portfolios in the USA) [23]. A Mean-variance-skewness-kurtosis entropy (MVSKEM) portfolio model was developed. An entropy component was included to improve diversification of the model. The capability of two different entropies; Gini-Simpson's and Shannon's were tested. A PGP was used to address different conflicting objectives, where the expected return, skewness, and entropy were maximized while the variance and kurtosis were minimized. In the first stage, each objective in the portfolio model was tackled, one by one, to achieve its desired level. In the second stage, deviations of these objectives were minimized. The two entropy measures allowed effective diversification of the portfolio without greatly degrading other objective values. The performance of Gini-Simpson's entropy was slightly better than Shannon's entropy in terms of Modified Sharpe Ratio (MSR) while the latter offered more diversity compared to the former.

3) FUZZY GOAL PROGRAMMING

De et al. proposed a fuzzy goal programming (FGP) approach to solve a multi-objective POP (National Stock Exchange (NSE) dataset) [24]. Return, risk, and liquidity constraints were used to formulate a portfolio model. Semi-absolute deviation was used to measure risk while triangular and trapezoidal membership functions were used to describe fuzzy return and fuzzy liquidity. A competitive-cum-compensatory operator was used to aggregate all the objectives of the portfolio model. Investors could set the preferred values of these objectives according to their needs. The proposed approach generated efficient portfolios for numerous strategies (changing different values of the objectives) in a competitive-cum-compensatory decision environment.

4) QUADRATIC PROGRAMMING

Puerto et al. proposed a scenario filtering approach based on Quadratic Programming to solve a mean-variance portfolio model (data from real-world stock markets: Dow Jones Industrial Average (DJIA), EUROSTOXX50, Financial Times Stock Exchange (FTSE100), and Standard & Poort's (S&P 500)) [18]. A Mixed Integer Quadratic Programming (MIQP) model was developed. A filtering technique was applied to the historical assets' return series to remove outliers and reduce estimation errors. The Mean-Variance portfolio model was used to select a portfolio, utilising the XPRESS solver. The MIQP model could solve small and medium size datasets efficiently.

C. MATHEURISTIC

Akbay et al. proposed a two-stage solution methodology that integrated Variable Neighbourhood Search (VNS) and Quadratic Programming (QP) to address the OR-Library benchmark dataset [39]. A cardinality constrained portfolio optimisation (CCPO) problem was formulated. In stage 1, a VNS was used to decide the combination of assets in a portfolio. In stage 2, a QP was employed to compute the proportion of assets in the portfolio. An initial solution and an asset selection pool were developed. Shaking, local search, and QP were iteratively performed to search for an optimal portfolio. A parallelization strategy was used to decrease computational time. The proposed methodology was highly competitive with the state-of-the-art algorithms [1], [65], [66], [67], [71], [72].

D. HYBRID APPROACHES

Yaman and Dalkılıç hybridised a Nonlinear Neural Network and a Genetic Algorithm (NNN-GA) to address the Mean-Variance Cardinality Constraint POP (MVCCPOP) (Istanbul Stock Exchange-30 (ISE-30) data) [52]. The POP was formulated as a Mixed Integer Quadratic Programming Problem (MIQP). Cardinality and bound constraints were considered as real-life constraints. A proportion of assets was represented by a chromosome in the GA. A Sharpe ratio of the portfolio model was maximized, while respecting cardinality. An NNN was then used to further minimize the risk of the portfolio model by solving primal and dual problems simultaneously. The proposed approach was compared with Active Set Method (ASM) and NNN. The return of the hybrid approach was comparable with, and higher than that of NNN and ASM respectively.

Mendonça et al. integrated an improved Non-dominated Sorting Genetic Algorithm II (NSGA-II) with a Neural Network Decision Maker 2 (NNDM-2) to tackle a multi-objective mean-CVaR portfolio optimisation model (benchmark dataset and Brazilian stock market dataset) [36]. The problem was formulated as a non-linear integer optimisation model, considering cardinality constraints and rebalancing. NSGA-II was used to address the portfolio model while NNDM-2 was used to approximate investor behaviour based on three functions (conservative, moderate and aggressive). An initial population was generated by randomly allocating the cardinality (number of assets) of each individual. A selection operator (non-dominated sorting and crowding distance) was used to select the best individuals. The selected individuals were used to form the population of the next generation. Two individuals of the new population were randomly chosen for a crossover operation. A mutation operator was used to change the recombined individuals. NNDM-2 with Multilayer Perceptron (MLP) was then used to guide investors in choosing the best solution among a set of Pareto frontier solutions produced by NSGA-II, based on their preference. Investors of all profiles generated significant profits. Aggresive investors gained more than the conservative ones.

Chen et al. integrated a two-stage clustering, a radial basis function (RBF) neural network, and a Genetic Algorithm (GA) to solve a mean-variance-skewness-kurtosis portfolio optimisation model (Shanghai Stock Exchange (SSE) dataset) [37]. A POP was formulated as a constrained non-linear programming model. A two-stage clustering was used to select assets for a diversified portfolio. All stocks were divided into seven clusters, where the centre of these clusters was chosen to construct the portfolio. A RBF neural network was applied to estimate the future returns of the selected assets by using mean square error (MSE) as the evaluation criteria. The trained network (met forecasting accuracy) was used to predict the daily return rates of the assets. Based on the prediction results of the RBF neural network, a GA was employed to search for the optimal portfolios under different risk preferences. The GA outperformed Simulated Annealing (SA) and mixed penalty function method (MPFM) in terms of yield (average return, Sharpe ratio) and risk management (system risk).

Leow et al. presented a hybrid approach by combining machine learning techniques and a metaheuristic algorithm to improve portfolio performance of a robo-advisor (web application) (United States Exchange Traded Funds (ETF)) [51]. By integrating traditional portfolio models with Twitter sentiments, two new models; Sentimental All-Weather (SAW) portfolio and Sentimental Modern Portfolio Theory (SMPT) were proposed. Two machine learning techniques, Valence Aware Dictionary and sEntiment Reasoner (VADER) and Google's Bidirectional Transformer for Financial Data (FinBERT) were used to capture up-to-date market conditions by converting tweets from various accounts into sentiments. A Genetic Algorithm (GA) was then used to optimise the models over the training period for multiple runs, with weekly rebalancing, by maximising cumulative returns and Sharpe Ratio and minimizing volatility. The proposed models outperformed buy-and-hold SPY index, MPT model, and Constant Rebalancing (CRB) model for an All-Weather portfolio model (in terms of Sharpe ratio, cumulative returns, and value-at-risk).

Simos et al. integrated power-activation feed-forward neuronet (PFN) and a two-stage Beetle Antennae Search (BAS) algorithm to tackle Time-varying Black Litterman Portfolio Optimisation (TV-BLPO) problem (real-world data from Yahoo Finance) [50]. By integrating the cardinality constraint and transaction cost, a POP was formulated as a non-linear programming (NLP) model. The investor's view in TV-BLPOP (a forecasting problem) was tackled using PFN in a speedy weights-and-structure-determination (WASD) algorithm. The optimal neuron (hidden-layer) weights and number of input-layer, of PFN were determined (MATLAB functions) to ensure the effectiveness of the neurons. The forecasting performance of PFN was evaluated by utilising symmetric mean absolute percentage error (SMAPE). The POP was splitted into two subproblems; binary integer programming problem (solved by binary BAS (BBAS)) and non-linear programming problem (solved by a modified BAS). BAS performed better than Artificial Bee Colony (ABC), Slime Mould Algorithm (SMA), and Differential Evolution (DE) in addresing the POP, especially for instances with large portfolio dimension (90 stocks).

Kizys et al. proposed a simulation-optimisation approach by integrating Variable Neighbourhood Search (VNS) metaheuristic with Monte Carlo simulation in addressing the benchmark OR-Library dataset [19]. A mathematical formulation based on stochastic portfolio model was formulated, where risk was treated as an objective function while return was treated as a constraint. A VNS was used to solve the portfolio model. The VNS was guided by a simulation technique which was utilised to predict the risk of each portfolio under uncertain circumstances. The advantages of the proposed approach were proven by various computational experiments.

Kara et al. combined ARMA-GARCH (1, 1) econometric model and Support Vector Regression (SVR) to tackle the POP (data from Istanbul Stock Exchange (BIST-30 index) and US Stock Market (Dow Jones index)) [31]. A Black-Litterman (BL) portfolio model was formulated. The proposed approach was divided into three stages. In stage 1, a ARMA-GARCH (1, 1) econometric modeling was used to obtain daily forecasts of 8 stock indicators. In stage 2, SVR was used to generate stock returns from the predicted indicators. In stage 3, the return forecasts were used as investor views to create a portfolio model. Results showed that the return of the hybrid approach was higher than the average returns of both indexes.

Bavarsad Salehpoor and Molla-Alizadeh-Zavardehi proposed a hybrid of metaheuristic algorithms to address the POP (data from Tehran stock market) [32]. Cardinality constrained portfolio optimisation models based on mean-variance, mean absolute deviation (MAD), semi variance (SV) and variance with skewness (VWS) were proposed. The authors hybridised five metaheuristic algorithms: Particle Swarm Optimisation (PSO), Simulated Annealing (SA), Genetic Algorithm (GA), Genetic Network Programming (GNP), and Electromagnetism-like algorithm (EM), with a diversification mechanism to enhance diversity and overcome local optimality. This mechanism was applied when the best objective function stopped changing after pre-specified iterations. The hybridized GA, GNP, and SA showed relatively good performance among all the hybridized algorithms.

Chen et al. hybridised improved firefly algorithm (IFA), eXtreme Gradient Boosting (XGBoost) and Monte Carlo method to address a real-world POP (Shanghai Stock Exchange (SSE) data) [53]. In Stage 1 (stock prediction), IFA and XGBoost were integrated to predict the prices of all the candidate assets for the next period. IFA was used to optimize the hyperparameters of XGBoost in a training process that use Mean square error (MSE) as the fitness function. This IFAXGBoost model was employed once the optimal hyperparameters were obtained. In Stage 2 (portfolio selection), a Mean-Variance (MV) model was used to determine the proportion of assets in a portfolio based on selected high-quality assets (high potential returns). A Monte Carlo method was used to randomly generate portfolios with different weight allocations. The mean return and variance of each portfolio were calculated. Portfolio with highest Sharpe ratio represented the best portfolio. Results showed that the hybrid approach outperformed the traditional method (without stock prediction) and benchmarks (IFAXGBoost + equal-weight portfolio model (1/N), machine learning + MV or 1/N, and Random + MV or 1/N) in terms of returns, risks, and return-risk ratio.

Chen et al. proposed a hybrid metaheuristic by integrating a firefly algorithm (FA) and a simulated annealing (SA) algorithm to solve a fuzzy portfolio model (hypothetical dataset) [25]. A possibilistic mean-semi-absolute deviation portfolio model was formulated. The FA algorithm was utilised as an exploitation process while the SA algorithm was used as an exploration process. A parameter "abandonment threshold" (AT) representing the maximum number of failed attempts of the FA algorithm was used to improve the global search ability of the FA algorithm. The SA algorithm was initiated when a potential solution was not being improved during the AT iterations. However, the exploration phase was not needed if the FA algorithm was able to discover the right part of the search space. The hybrid approach outperformed FA, SA, Particle Swarm Optimisation, and Genetic Algorithm (GA) in finding optimal portfolios. The hybrid FA-SA converged the fastest among all the algorithms.

Kalayci et al. applied a hybrid algorithm that integrated important elements from various algorithms such as Continuous Ant Colony Optimisation (CACO), Genetic Algorithm (GA) and Artificial Bee Colony (ABC) in addressing seven publicly available benchmark problems (five from OR-Library and one from XU030, and XU100 indices each) [38]. A mathematical formulation of cardinality constrained portfolio optimisation (CCPO) problem was presented. Each algorithm contributed an important role in the hybrid method. A sample gaussian formulation was applied in the CACO algorithm to iteratively improve and update an archive of solutions. The GA algorithm employed an elitism mechanism to; transfer best solutions from the previous generation to the next generation and avoid rediscovering solutions that were found in the previous generations. A modification parameter was utilized in the ABC algorithm to randomly restrict changes made to a solution. This prevented the solution from being over-diversified in the solution space. The hybrid algorithms was competitive with the state-of-the-art algorithms in the literature.

Mehlawat et al. hybridized a Non-dominated Sorting Genetic Algorithm II (NSGA-II) and a Genetic Algorithm (GA) to address the POP (NASDAQ-100 index) [54]. A mean–mean absolute semi-deviation–skewness portfolio model was formulated. In stage 1, NSGA-II was used to address cardinality and bound constraints in the portfolio model to obtain a set of efficient solutions. If an investor was satisfied with one of the obtained results, the process stopped at this stage. Otherwise, risk tolerance and skewness were collected from the investor and included in the portfolio model. In stage 2, the non-dominated solutions from first stage was used by GA as starting population of solutions. Roulette wheel selection operator and simulated binary crossover were used to produce offspring for the next generation. A non-uniform mutation operator was used to produce new solutions in the population. The proposed approach could produce more reliable results than that of [73].

E. MACHINE LEARNING

Wang et al. applied long short-term memory (LSTM) networks to solve a real-world POP (UK Stock Exchange 100 index (FTSE 100)) [40]. In stage 1, 21 assets were randomly chosen from FTSE 100. The long short-term memory networks (LSTM) were used to predict the return of these assets. The assets were then ranked based on their predicted returns. An asset with a higher return was given a higher ranking than an asset with a lower return. A random search method was performed to optimize the hyperparameters in LSTM. In stage 2, the proportion of each asset was determined using a Mean-Variance (MV) model. A function was developed using Python to randomly generate 50,000 different portfolios with different weight allocation. The portfolio with the lowest variance was selected as the best portfolio. Different number of assets (4, 5, 6, 7, 8, 9, 10) were tested. Results showed that portfolio that held 10 assets performed the best in terms of highest mean return, Sortino ratio, Sharpe ratio, and lowest standard deviation.

Paiva et al. employed a Support Vector Machine (SVM) method to solve a real-world POP (Sao Paulo Stock Exchange Index (Ibovespa)) [33]. The SVM was used to classify assets. Only assets that have potential to achieve a given daily return were considered. Mean-Variance (MV) model was then used to decide the amount of capital distributed to each asset, where the assets with lower variance were preferred. The proposed model was called SVM + MV. Results were compared with SVM + 1/N and Random + MV models. The proposed model generated the highest daily return average.

Aboussalah and Lee proposed a Stacked Deep Dynamic Recurrent Reinforcement Learning (SDDRRL) architecture to solve real-time POP (S&P 500 index) [41]. Stocks were chosen from different sectors to promote portfolio diversification. Recent market conditions were captured by SDDRRL. A portfolio model was rebalanced if needed. An automated Gaussian Process (GP) with Expected Improvement (EI) was used as an acquisition function in SDDRRL to determine the best possible architecture topology. The total return of portfolio was maximized while satisfying cardinality constraints. Results showed that the proposed approach outperformed the Mean-Variance (MV) model, risk parity model, and the uniform buy-and-hold (UBAH) index in terms of total return.

Ma et al. proposed machine learning and deep learning models to tackle a real-world POP (data from China Securities 100 Index). Random forest (RF) and Support Vector Regression (SVR) were chosen for the machine learning models while Long Short Term Memory (LSTM) neural network, Deep Multilayer Perceptron (DMLP), and Convolutional neural network (CNN) were included in the deep learning models. Grid research was used to discover the optimal hyperparameter of all these models [55]. These models were utilised to predict return of the assets (stock preselection). The predicted results were then integrated into Mean-Variance and Omega portfolio models in constructing portfolio models.

V. RESEARCH LIMITATIONS (CHALLENGES)

NABAS performed comparably to PSO and GA and was more efficient in terms of computational cost than both algorithms [15]. However, NABAS's performance deteriorates as the size of portfolio increases. It could be due to its nature which lacks exploration capability as a single-solution based metaheuristic. Population-based metaheuristics are believed to be better in this regard. Furthermore, parameter settings were required for metaheuristics to balance between exploration and exploitation rate in ensuring the efficiency of these algorithms [44], [48].

Mathematical optimisation (MO) could solve small and medium sized POP efficiently provided sufficient time and memory are given. But, MO were struggling to solve large sized POP [18], [30].

Hybridisation is one of the most popular approaches among researchers in addressing the POP. The idea is to combine strengths and avoid weaknesses of different approaches. Nevertheless, implementing an effective hybrid method (especially a closely integrated hybrid) is challenging as it is hard to attain the desired synergy among the algorithms. In addition, a comprehensive analysis is required to determine the most suitable parameter values of various algorithms to achieve optimal performance [38].

Among the limitations reported were; the proposed methodology were compared to limited methodologies [15], the proposed methodology were only tested on a particular stock market [33], [40], [47], [53], the proposed methodology were lacking of input features to predict returns [40], [55], the proposed model were tested on limited methodologies [26], [43], [53] and lacking of asset selection in constructing the asset universe for the models [45].

Collecting accurate and reliable data could also be challenging. Historical data was frequently used to predict the return of portfolio optimization models. This data might be insufficient or low quality, resulting in inaccurate predictions. In [17], only 74 assets in CSI Index (Shanghai and Shenzhen Stock Markets) were considered. The data incompleteness was primarily due to the unlisting or delisting of corresponding companies during the inspection period.

Unlike the benchmark POP (fixed constraints), real-world POP often have their own specific constraints to fulfill, that could limit the practical usefulness of the proposed model. Total transaction costs and tax-liability constraints were included in [15]. Minimum transaction lots and dynamic risk preference constraints were involved in [43]. Cardinality, floor and ceiling, and transaction costs constraints were integrated to the problem [45]. Transaction consists and bankruptcy constraints were included in [46]. Turnover, floor and ceiling, and cardinality constraints were integrated to the problem [29]. Cardinality, contingent, and floor and ceiling constraints were included in the problem [30]. Cardinality, floor and ceiling, contingent, and transaction costs constraints were included in the problem [21].

Adopting portfolio optimization models could be difficult, particularly in real-world situations with large portfolios and multiple assets. In [50], the proposed methodology was only tested up to 90 stocks with real-world data. In [43], only 6 stocks (first example) and 20 stocks (second example) from Shanghai Stock Exchange were utilized to examine the proposed methodology. In [46], only 5 assets was used to illustrate the efficiency of the proposed methodology. In [34], the proposed methodology was only tested up to 40 stocks.

In general, portfolio optimization models were influenced by market volatility. The financial market was inherently unstable and unpredictable, frequently changing and evolving. It could be impacted by economic conditions, news, government policies, political events, interest rates, and so on. The high degree of uncertainty and volatility increased the challenges to maintain an optimal portfolio. Next, selecting the best model for a particular POP could be difficult. There were numerous portfolio models available, each with its advantage and disadvantage. Comprehensive knowledge of financial markets and investment techniques were required to make a correct decision.

VI. CLASSIFICATION OF PROBLEM INSTANCES

There are several datasets used by researchers to assess the capability of various algorithms and portfolio models. Table 3 shows the classification of the datasets. The datasets can be classified into three: benchmark, real-world, and hypothetical (based on situations or ideas that are possible and imagined).

A. BENCHMARK DATASETS

Table 4 shows the description for the benchmark datasets. They are commonly used to compare algorithms in an objective manner. The OR-Library dataset is the most favoured benchmark among researchers, followed by XU030 and XU100 indices, and Johannesburg Stock Exchange.

Figure 4 displays the solution methodologies for the POP arranged according to datasets. The state-of-the art methods for the OR-Library dataset are; a hybrid of Variable Neighbourhood Search and Monte Carlo Search from [19], Modified Squirrel Search Algorithm from [42], and Artificial Bee Colony from [49]. Population-based metaheuristic is a favoured approach in addressing the OR-Library datasets.

Performance indicators were used to compare the capability of algorithms. They can be categorized into three categories: diversity-based, convergence-based, and hybridization-based. Table 5 shows the performance metrics used for the benchmark datasets. Median Percentage Error (MEDPE), Mean Percentage Error (MPE), and Variance of Return Error (VRE) were the most commonly used metrics. Diversity Metric (Δ), Epsilon indicator ($I\epsilon$), Error Ratio (Er), Spacing metric (S), and Inverted generational distance (IGD) were the least commonly used metrics.

B. REAL-WORLD DATASETS

Table 6 exhibits the real-world datasets. There are more real-world datasets than benchmark datasets. Researchers often used real-world datasets from various stock markets to assess the performance of their proposed portfolio models.

C. HYPOTHETICAL DATASETS

Table 7 shows the three hypothetical datasets that are found in the literature. Various numerical experiments were performed by making assumption of investment scenarios.

VII. RESEARCH OPPORTUNITIES

Some researchers wished to extend their models by integrating real-life constraints in an effort to generate more practical portfolios. For example, Khan et al. planned to introduce cardinality, liquidity, and regulatory constraints to their

TABLE 3. Classification of datasets.

Year	Dataset	Benchmark (B) / Real-world (RW) / Hypothetical (H)	Methodology	Solution Category	Reference
2018	OR	B	MOEA	P-MH	[20]
	CSM	RW	MOPSO	P-MH	[63]
	FFL	RW	MOEA	P-MH	[21]
	FFL	RW	MILP	MO	[22]
	ISE30, FFL	RW	PGP	MO	[23]
	NSE	RW	FGP	MO	[24]
	-	H	FA, SA	HA	[25]
2019	OR	B	ARMOPSO	P-MH	[27]
	S&P500	RW	NSGA-II	P-MH	[26]
	DJIA, FF49, NASDAQ	RW	IMOPSO	P-MH	[28]
	FTSE100	RW	Info Based EA	P-MH	[29]
	BIST30, DJI	RW	ARMA-GARCH, SVR	HA	[31]
	TSM	RW	EM, GA, GNP, PSO, SA (hybridised with diversification mechanism)	HA	[32]
	NSE	RW	PGP	MO	[30]
	BSM	RW	SVM	ML	[33]
2020	OR	B	VNS, QP	M	[39]
	OR, XU030, XU100	B	CACO, GA, ABC	HA	[38]
	BSM	RW	NSGA, NNDM-2	HA	[36]
	SSE	RW	TSC, RBFNN, GA	HA	[37]
	NYSE	RW	NSGA-II	P-MH	[34]
	NSM	RW	BAS	P-MH	[35]
	FTSE100	RW	LSTM	ML	[40]
	S&P500	RW	SDDRRL	ML	[41]
2021	OR, JSE	B	MSSA	P-MH	[42]
	OR, XU030, XU100	B	ABC	P-MH	[49]
	United States ETFs	RW	VADER, finBERT, GA	HA	[51]
	ISE-30	RW	GA, NNN	HA	[52]
	SSE	RW	XGBoost, IFA, MCM	HA	[53]
	NSM	RW	NSGA-II, GA	HA	[54]
	SSE	RW	GA	P-MH	[43]
	NSE, S&P 500, NSM	RW	GA	P-MH	[45]
	BSM	RW	NSGA-II	P-MH	[47]
	HKSM	RW	LQPSO	P-MH	[48]
	SSSM	RW	RF, SVR, LSTM, DMLP, CNN	ML	[55]
	Data from Yahoo Finance	RW	PFN, BAS	HA	[50]
	Data obtained from [57]	H	GA	P-MH	[44]
	-	H	GA	P-MH	[46]
2022	OR	B	VNS, MCS	HA	[19]
	NSE	RW	NABAS	P-MH	[15]
	NYSE, HKSM	RW	MGO	P-MH	[16]
	SSSM, SSE	RW	IGA	P-MH	[17]
	NYSE, FTSE100, EUROSTOXX50, S&P500	RW	QP	MO	[18]

model [15]. Zhang et al. [16] and Chen at al. [25] would like to add cardinality and minimum transaction lot constraints to the problem. Dhaini and Mansour wanted to consider transaction costs and pre-assignment constraints in the addressing the POP [42]. Kara et al. hoped to include round lots, floor and ceiling, cardinality, class, and pre-assignment constraints to the problem [31]. Meghwani and Thakur planned to include chance, class, and round-lot constraints to their proposed model [21]. Gupta et al. suggested to include the liquidity of assets along with the expected return in their portfolio optimisation model [45]. Barroso et al. hoped to consider the social responsibility factor when making decision for investments [47]. Kalayci et al. [38] and Cura [49] intended to include minimum transaction lots and transaction costs to the problem. Gong et al. suggested to incorporate more objectives into the proposed model, such as portfolio efficiency and liquidity [17].

Our survey reveals that most work on POP are based on real-world datasets. This is not surprising as there are limited numbers of publicly available benchmark datasets for researchers to focus on. Therefore, an extension to the pool of benchmark datasets is highly recommended. It is suggested that benchmark datasets that consider real-world constraints and imitate real-world POP to be introduced in future. This allows researchers to compare their algorithms and models objectively and practically.

TABLE 4. Description for the benchmark datasets.

Dataset	Description	Reference
OR-Library (OR) [1], [74]	- Weekly price data of 7 stock markets; 31 assets from Hang Seng (Hong Kong), 85 assets from DAX100 (Germany), 89 assets from FTSE100 (UK), 98 assets from S&P100 (USA), 225 assets from Nikkei225 (Japan), 457 assets from S&P500 (USA), and 1317 assets from Russell2000 (USA) from March 1992 to September 1997 [74].	[20], [38], [39], [49]
	- Weekly price data of 5 stock markets (31 assets from Hang Seng (Hong Kong), 85 assets from DAX100 (Germany), 89 assets from FTSE100 (UK), 98 assets from S&P100 (USA), 225 assets from Nikkei225 (Japan)) from March 1992 to September 1997 [1].	[19], [27], [42]
XU030 and XU100 indices [72]	- Daily prices of 30 (XU030) and 99 (XU100) assets from May 2013 to April 2016.	[38], [49]
Johannesburg Stock Exchange (JSE) [75]	- Top 100 stocks by market capitalization on JSE [76].	[42]

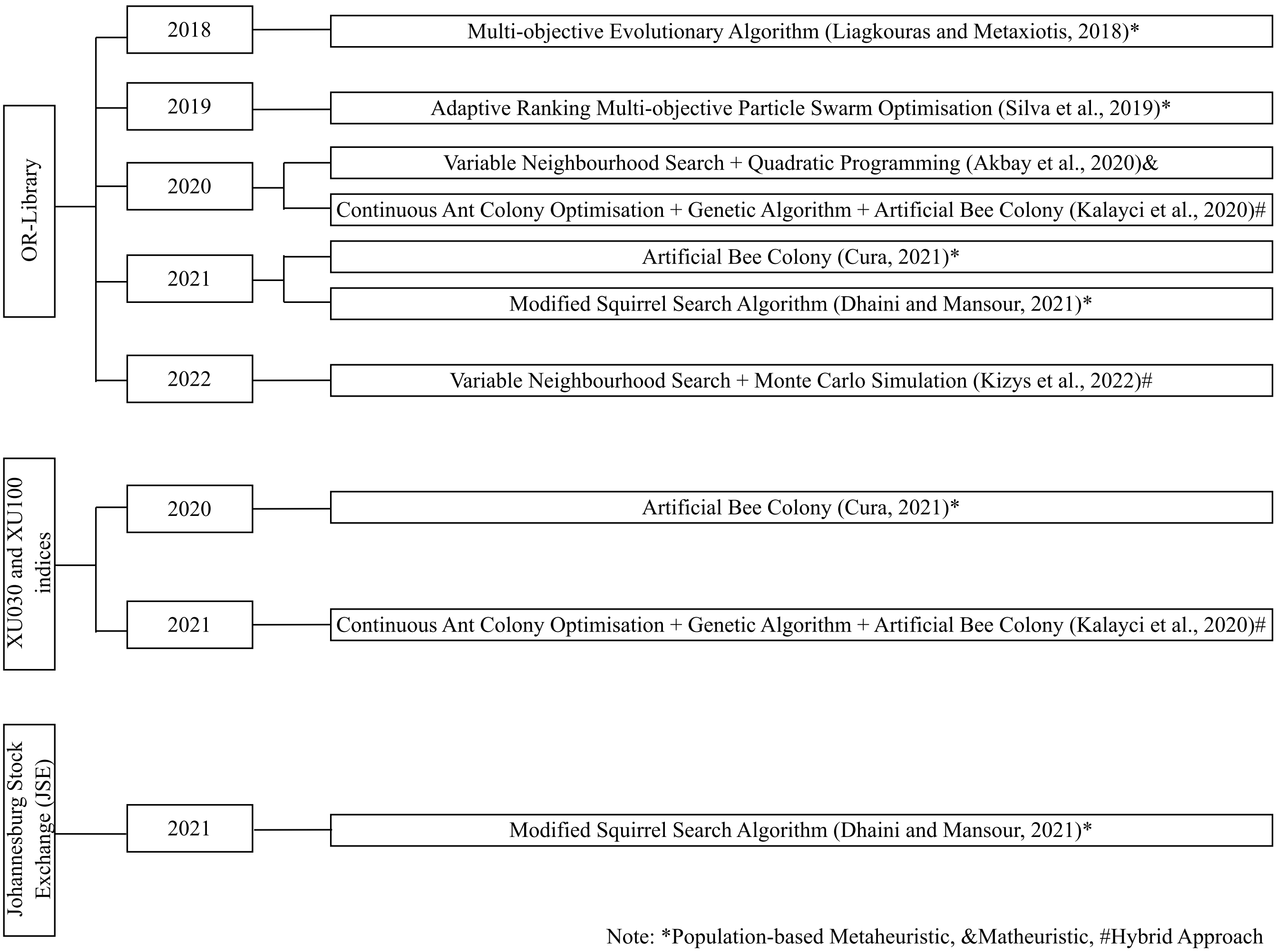


FIGURE 4. Solution methodologies for the POP (According to datasets).

Some researchers hoped to test their methodologies and models on POP with larger portfolio size. Khan et al. planned to provide more options for investors to invest by expanding the portfolio size to over 100 companies [15]. Simos et al. wanted to examine the proposed Beetle Antennae Search (BAS) algorithm in a larger portfolio environment [50].

TABLE 5. Performance metrics for benchmark datasets.

Category	Performance Metrics	Description	Reference
Convergence-based Metrics	Epsilon indicator (Iϵ) [77]	A binary indicator that shows the degree to which one approximation set is worse than another when all objectives are considered.	[20]
	Error Ratio (Er) [78]	Percentage of solutions that are not in the optimal efficient frontier.	[27]
	Generational Distance/ Gravitational Distance (GD) [78], [79]	Distance between obtained efficient frontier from optimal frontier.	[27], [42]
	Maximum Percentage Error (MAXPE) [1]	Maximum value of all the percentage deviations.	[27], [38], [39], [49]
	Mean Euclidean Distance (MEUCD) [71]	Average straight line distance between each portfolio alternatives of the optimal efficient frontier and the obtained efficient frontier.	[38], [39], [49]
	Mean Percentage Error (MPE) [1]	Average percentage deviations of the obtained efficient frontier from the optimal efficient frontier.	[27], [38], [39], [42], [49]
	Mean Return Error (MRE) [71]	Average value of all the errors (vertical distance corresponding to a certain level of risk).	[27], [38], [42], [49]
	Median Percentage Error (MEDPE) [1]	Median value of all the percentage deviations.	[27], [38], [39], [42], [49]
	Minimum Percentage Error (MINPE) [1]	Minimum value of all the percentage deviations.	[27], [38], [39], [49]
	Variance of Return Error (VRE) [71]	Average value of all the errors (horizontal distance corresponding to a certain level of return).	[27], [38], [39], [42], [49]
Diversity-based Metrics	Spacing Metric (S)	Dispersion between the optimal efficient frontier and the set of non-dominated solutions.	[27]
	Diversity Metric (Δ) [80]	Dispersion of the solution set.	[27]
Hybridization-based Metrics	Hypervolume Indicator (HV)	Proximity between the optimal efficient frontier and obtained efficient frontier, and also the diversity of the obtained solutions.	[20], [27]
	Inverted Generational Distance (IGD)	Reverse order of generational distance.	[20]

Dai and Qin considered testing their proposed methodology in larger-scale numerical experiments to strengthen the persuasiveness of their theoretical model [43].

Many researchers provided suggestions on how to improve the performance of their proposed algorithms. Khan et al. believed that the efficiency and convergence rate of their NABAS algorithm could be further improved [15]. Kizys et al. hoped to hybridize a statistical learning technique with a metaheuristic-based approach (known as learnheuristics) for the dynamic stochastic POP [19]. Dhaini and Mansour suggested to hybridize their MSSA algorithm with features from other metaheuristic techniques to improve its exploration and exploitation capabilities [42]. Yaman and Dalkılıç hypothesized that inclusion of fuzzy logic will improve the performance of their proposed method (NNNGA) [52]. Cura intended to focus on the parallel version of their proposed approach (ABC) to enhance its computational efficiency [49]. Liagkouras and Metaxiotis suggested to substitute the crossover and mutation selection probabilities with a variable that will be updated at run-time according to the performance of their algorithm [20].

Some researchers hoped to extend their algorithms to tackle multi-period POP (where investors are required to review and revise their portfolio in each period). Kizys et al. were looking forward to expand the proposed approach to address a multi-period stochastic POP that allows investors to rebalance their assets in the portfolio [19]. A few researchers intended to test their proposed methodologies on multiple investment horizons [25], [38], [39], [54].

Several researchers planned to consider higher order moments (mean (first order), variance (second order), kurtosis (third order) skewness (fourth order)) of portfolio returns in the POP. In addition to mean and standard deviation, Kizys et al. intended to integrate higher order moments to portfolio returns [19]. Puerto et al. wanted to consider higher order filtered skewness and kurtosis in addressing the problem [18]. Huang et al. [44] and Chen at al. [53] wished to investigate the influence of skewness and kurtosis to the problem.

Some researchers intended to apply different risk measures to the problem. Kizys et al. wanted to consider Value at Risk (VaR) and Conditional Value at Risk (CVaR) in POP [19]. Huang et al. intended to study the effect of semi variance

TABLE 6. Real-world datasets.

Dataset	Description	Reference
Chinese Stock Market (CSM)	- Daily prices of 10 stocks from January 1, 2006 to December 31, 2010. - The dataset is divided into four observations periods; entire period, first sub-period from January 1, 2006 to October 31, 2007 (bull market), second sub-period from November 1, 2007 to October 31, 2008 (bear market), and last sub-period from November 1, 2008 to December 31, 2010 (steady market).	[63]
EUROSTOXX50 index	- Represents leading blue-chip index in Europe. - Weekly price data of 49 assets from April 22, 2002 to April 4, 2016.	[18]
BIST-30 Index	- One of the indexes from Turkish Stock Exchange/ Istanbul Stock Exchange. - Data from 1996 to 2016. - Stock returns collected from Bloomberg data terminal.	[31]
Fama and French Data Library (FFL)/ Kenneth R. French's website [1]	- Represent the distinct industry portfolios of 38 and 48 industries (FF38 and FF48). - 1151 days daily returns data from January 3, 2012 to July 29, 2016.	[21]
	- Daily returns data on 100 portfolios formed on size and book-to-market (10 x 10).	[22]
	- Monthly returns data of 12 industry portfolios from January 2005 to November 2016 in the United States of America.	[23]
	- 49 assets from FF49 industries from July 1969 to July 2015.	[28]
Brazilian Stock Market (BSM)	- 53 assets from January 2011 to December 2015.	[36]
	- 53 assets from January 2010 to December 2015.	[47]
	- São Paulo Stock Exchange Index (Ibovespa) from June 2001 to December 2006.	[33]
Istanbul Stock Exchange (ISE) 30 [2, 3]	- Monthly returns of 21 assets from January 2005 to November 2016.	[23]
	- Daily prices of assets from June 2015 to May 2017.	[52]
National Stock Exchange (NSE) [4]	- Three sets of stocks from three different indicies (NIFTY 50, NIFTY Smallcap 100, and NIFTY 500) from October 7, 2016 to July 23, 2018. - Data collected from PROWESS.	[24]
	- Includes historical data of 20 assets.	[30]
	- 4-year daily return data of 20 assets from January 1, 2016 to December 31, 2019.	[45]
Standard & Poor's (S&P) 500 index	- Daily close prices of assets from [81].	[26]
	- 10 stocks from January 1, 2013 to July 31st, 2017.	[41]
	- 8-year daily return data from January 1, 2011 to December 31, 2018.	[45]
	- Weekly price data of 442 assets from April 18, 2005 to April 4, 2016.	[18]
Financial Times Stock Exchange (FTSE) 100 index	- 21 daily stock data from March 1994 to March 2019.	[40]
	- Provided by [82].	[29]
	- Weekly price data of 83 assets from April 19, 2004 to April 4, 2016 are collected.	[18]
Tehran Stock Market (TSM)	- Historical daily data of 50 companies.	[32]
NASDAQ Stock Market (NSM) [5]	- 8-year daily return data from January 1, 2011 to December 31, 2018.	[45]
	- Stock data of 20 companies.	[15]
	- 8 year daily data from January 1, 2011 to December 31, 2018.	[54]
	- 50 stock price data in 2017.	[35]
	- 82 assets from NASDAQ100 from November 2004 to April 2016.	[28]
Shanghai and Shenzhen Stock Market (SSSM)	- 49 stocks from China Securities Index (CSI) 100 from January 4, 2007 to December 31, 2015.	[55]
	- 74 assets of CSI 100 from January 4, 2011 to January 4, 2016.	[17]
New York Stock Exchange (NYSE)	- Dow Jones Index (DJI) index from 1978 to 2018. - Stock returns collected from Bloomberg data terminal.	[31]
	- Weekly returns of stocks from January 2015 to January 2018.	[34]
	- 27 stocks from DJI index. - Data from January 2018 to December 2019.	[16]
	- 28 assets from DJI. - Weekly price data of assets from May 7, 1990 to April 4, 2016.	[18]
	- 28 assets from DJI index from February 1990 to April 2016.	[28]
Shanghai Stock Exchange (SSE)	- Daily trading data of SSE 50 index from January 4, 2010 to February 20, 2017. - Provided by CSMAR China Stock Market Trading Database. - SSE 50 index evaluates the stocks on SSE according to the total market value and turnover to choose the best 50 as the index's constituents (representsthe most representative stocks in terms of transaction size and liquidity in market)	[37]
	- 6 stocks are chosen for the first experiment while 20 stocks are chosen for the second experiment.	[43]
	- 24 stocks of SSE50 index from November 2009 to November 2019.	[53]
	- 101 stocks of SSE 180 index from December 24, 2010 to December 30, 2016.	[17]
United States Exchange Traded Funds (ETFs)	- One set of ETFs for sector-based diversification and another set for asset-based diversification.	[51]
Hong Kong Stock Market (HKSM)	- 20 stocks from June 2017 to August 2018.	[48]
	- 40 stocks from Hang Seng Index (HSI). - Data from January 2018 to December 2019.	[16]

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html Last accessed: November 1, 2022.
[2] https://datastore.borsaistanbul.com/ Last accessed: November 1, 2022.
[3] https://tr.investing.com/ Last accessed: November 1, 2022.
[4] https://www.cmie.com/ Last accessed: November 1, 2022.
[5] https://www.kaggle.com/datasets/paultimothymooney/stock-market-data Last accessed: November 1, 2022.

TABLE 7. Hypothetical datasets.

Description	Reference
- Possibility distributions of returns for 9 stocks are provided (calculated based on sample percentile method used in [83])	[25]
- Uncertain return of 10 assets taken from [57] are provided.	[44]
- Fuzzy asset return rates of 5 assets are provided.	[46]

(SV) and Semi-absolute Deviation (SAD) [44]. Akbay et al. wanted to develop new structures for the problem by considering different risk measures like VaR, CVaR, and Mean Absolute Deviation (MAD) [39]. Chen et al. intended to consider all possible risk preference parameters when optimizing POP [37]. Chen at al. wished to include SV, VaR, and CVaR into their proposed model to reflect the real investment experience [53].

Various researchers hoped to compare their proposed methodology with other methodologies. Khan et al. wanted to extend the comparison of their NABAS algorithm with robust optimisation and stochastic programming [15]. Simos et al. wished to compare their proposed BAS methodology with other popular metaheuristic techniques [50]. Huang et al. intended to compare their GA algorithm with other heuristic algorithms [44]. Kaucic wished to study the performance of other bio-inspired algorithms and constraint handling techniques on their proposed model [28].

A number of researchers wanted to test their proposed approach on other portfolio models. Lu and He intended to utilize their LQPSO algorithm to address the dynamic mean-variance portfolio optimisation model [48]. Khan et al. suggested to investigate the effect of Sharpe ratio, on the performance of their BAS algorithm [35]. Dhaini and Mansour suggested to apply their MSSA algorithm to other portfolio models, such as inequality cardinality constrained, inequality cardinality pre-assignment constrained, and cardinality round-lot pre-assignment constrained portfolio optimisation problems [42].

Several researchers planned to work on new variants of POP. For example, Chen et al. hoped to investigate on dynamic higher-order-moment POP [37]. Gong et al. suggested to extend their proposed model to a credibilistic environment and investigate the performance of the model under dynamic-risk tolerance and expected-return levels [17]. Dai and Qin would consider a more general portfolio model that can describe both static and dynamic risk preference level [43].

Some researchers intended to investigate the applicability of their proposed methodology on other financial markets. Zhang et al. wished to implement their proposed trading system in futures and bond markets [16]. Gong et al. wanted to test their IGA algorithm on S&P 500 market data [17]. Akbay et al. hoped to test their proposed methodology (VNS + QP) on other datasets (markets) to evaluate how well it reacts to market specific variability [39].

Certain researchers think that inclusion of input indicators may further improve their proposed model or approach. Ma et al. wished to include more efficient input features such as news, economic indicators, technical indicators, and exchange rate to train predictive models and improve the performance of their proposed models for daily investment trading [55]. Barroso et al. wanted to consider other Technical Analysis indicators to validate their proposed approach [47]. Wang et al. intended to include other input indicators such as interest rates, government policies and public events to their proposed model [40].

Applying methodologies shown to be effective in other domains is a valuable research opportunity. In this survey, we notice the lack of work on hyper-heuristic in addresing the POP. Leow et al. suggested using rules or machine learning models to choose the best performers for each asset class in Exchange Traded Funds (ETF) selection [51]. Leow et al. intended to include other traditional algorithm such as Hierarchial Risk Parity(HRP) to improve the performance of their proposed portfolio model [51].

Some researchers intended to implement an appropriate selection of parameters to improve the model's performance. Mehlawat et al. wanted to perform a careful selection of the model parameters such as bounds on investment and level of diversification to improve the performance of the proposed model [54].

A researcher proposed to involve investors in the preliminary investigation of the stock markets to filter those assets which do not meet their requirements [45].

VIII. CONCLUSION

POP is an important research area and the number of publications is increasing every year. We reviewed the recent methodological articles between 2018 to 2022, with the aim to extend the taxonomy of the POP. We provide the definition, variants and constraints of the POP. We describe the operations and achievement of the solution methodologies for the benchmark, real-world, and hypothetical datasets. The solution methodologies are categorised. Benchmark repositories as well as the state-of-art methodologies are presented. In addition, we propose the emerging trends. Furthermore, we examine the research challenges and opportunities for the POP. We found that; population-based metaheuristics are the most preferred techniques among researchers in addressing the POP, hybrid approaches seem to be an emerging trend (2018 onwards) and OR-Library is the most popular benchmark dataset for researchers to compare their methodologies in addressing POP. We suppose that this survey paper contributes an useful insight to researchers.

REFERENCES

[1] T.-J. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha, ''Heuristics for cardinality constrained portfolio optimisation,'' *Comput. Oper. Res.*, vol. 27, no. 13, pp. 1271–1302, 2000.

[2] L. Guiso, P. Sapienza, and L. Zingales, ''Time varying risk aversion,'' *J. Financial Econ.*, vol. 128, no. 3, pp. 403–421, 2018.

[3] S. L. Goh, S. N. Sze, N. R. Sabar, S. Abdullah, and G. Kendall, ''A 2-stage approach for the nurse rostering problem,'' *IEEE Access*, vol. 10, pp. 69591–69604, 2022.

[4] C. M. Ngoo, S. L. Goh, S. N. Sze, N. R. Sabar, S. Abdullah, and G. Kendall, ''A survey of the nurse rostering solution methodologies: The state-of-the-art and emerging trends,'' *IEEE Access*, vol. 10, pp. 56504–56524, 2022.

[5] N. R. Sabar, S. L. Goh, A. Turky, and G. Kendall, ''Population-based iterated local search approach for dynamic vehicle routing problems,'' *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 2933–2943, Oct. 2022.

[6] H. M. Markowitz, *Portfolio Selection*. New Haven, CT, USA: Yale Univ. Press, 1968.

[7] P. Skolpadungket, K. Dahal, and N. Harnpornchai, ''Portfolio optimization using multi-objective genetic algorithms,'' in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2007, pp. 516–523.

[8] H. Konno and H. Yamazaki, ''Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market,'' *Manag. Sci.*, vol. 37, no. 5, pp. 519–531, 1991.

[9] R. T. Rockafellar and S. Uryasev, ''Optimization of conditional value-at-risk,'' *J. Risk*, vol. 2, pp. 21–42, Feb. 2000.

[10] M. R. Young, ''A minimax portfolio selection rule with linear programming solution,'' *Manage. Sci.*, vol. 44, no. 5, pp. 673–683, 1998.

[11] B. Aouni, C. Colapinto, and D. La Torre, ''Financial portfolio management through the goal programming model: Current state-of-the-art,'' *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 536–545, Apr. 2014.

[12] M. Masmoudi and F. B. Abdelaziz, ''Portfolio selection problem: A review of deterministic and stochastic multiple objective programming models,'' *Ann. Oper. Res.*, vol. 267, nos. 1–2, pp. 335–352, Aug. 2018.

[13] C. B. Kalayci, O. Ertenlice, and M. A. Akbay, ''A comprehensive review of deterministic models and applications for mean-variance portfolio optimization,'' *Expert Syst. Appl.*, vol. 125, pp. 345–368, Jul. 2019.

[14] K. Deb, *Multi-Objective Optimization*. Boston, MA, USA: Springer, 2005, pp. 273–316.

[15] A. T. Khan, X. Cao, I. Brajevic, P. S. Stanimirovic, V. N. Katsikis, and S. Li, ''Non-linear activated beetle antennae search: A novel technique for non-convex tax-aware portfolio optimization problem,'' *Expert Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116631.

[16] Y. Zhang, W. Liu, and X. Yang, ''An automatic trading system for fuzzy portfolio optimization problem with sell orders,'' *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115822.

[17] X. Gong, L. Min, and C. Yu, ''Multi-period portfolio selection under the coherent fuzzy environment with dynamic risk-tolerance and expected-return levels,'' *Appl. Soft Comput.*, vol. 114, Jan. 2022, Art. no. 108104.

[18] J. Puerto, F. Ricca, M. Rodríguez-Madrena, and A. Scozzari, ''A combinatorial optimization approach to scenario filtering in portfolio selection,'' *Comput. Oper. Res.*, vol. 142, Jun. 2022, Art. no. 105701.

[19] R. Kizys, J. Doering, A. A. Juan, O. Polat, L. Calvet, and J. Panadero, ''A simheuristic algorithm for the portfolio optimization problem with random returns and noisy covariances,'' *Comput. Oper. Res.*, vol. 139, Mar. 2022, Art. no. 105631.

[20] K. Liagkouras and K. Metaxiotis, ''A new efficiently encoded multiobjective algorithm for the solution of the cardinality constrained portfolio optimization problem,'' *Ann. Oper. Res.*, vol. 267, nos. 1–2, pp. 281–319, Aug. 2018.

[21] S. S. Meghwani and M. Thakur, ''Multi-objective heuristic algorithms for practical portfolio optimization and rebalancing with transaction cost,'' *Appl. Soft Comput.*, vol. 67, pp. 865–894, Jun. 2018.

[22] O. Babat, J. C. Vera, and L. F. Zuluaga, ''Computing near-optimal value-at-risk portfolios using integer programming techniques,'' *Eur. J. Oper. Res.*, vol. 266, no. 1, pp. 304–315, Apr. 2018.

[23] M. Aksaraylı and O. Pala, ''A polynomial goal programming model for portfolio optimization based on entropy and higher moments,'' *Expert Syst. Appl.*, vol. 94, pp. 185–192, Mar. 2018.

[24] M. De, B. K. Mangaraj, and K. B. Das, ''A fuzzy goal programming model in portfolio selection under competitive-cum-compensatory decision strategies,'' *Appl. Soft Comput.*, vol. 73, pp. 635–646, Dec. 2018.

[25] W. Chen, Y. Wang, and M. K. Mehlawat, ''A hybrid FA–SA algorithm for fuzzy portfolio selection with transaction costs,'' *Ann. Oper. Res.*, vol. 269, nos. 1–2, pp. 129–147, Oct. 2018.

[26] J. Goldkamp and M. Dehghanimohammadabadi, ''Evolutionary multi-objective optimization for multivariate pairs trading,'' *Expert Syst. Appl.*, vol. 135, pp. 113–128, Nov. 2019.

[27] Y. L. T. V. Silva, A. B. Herthel, and A. Subramanian, ''A multi-objective evolutionary algorithm for a class of mean-variance portfolio selection problems,'' *Expert Syst. Appl.*, vol. 133, pp. 225–241, Nov. 2019.

[28] M. Kaucic, ''Equity portfolio management with cardinality constraints and risk parity control using multi-objective particle swarm optimization,'' *Comput. Oper. Res.*, vol. 109, pp. 300–316, Sep. 2019.

[29] K. Liagkouras and K. Metaxiotis, ''Improving the performance of evolutionary algorithms: A new approach utilizing information from the evolutionary process and its application to the fuzzy portfolio optimization problem,'' *Ann. Oper. Res.*, vol. 272, nos. 1–2, pp. 119–137, Jan. 2019.

[30] P. Gupta, M. K. Mehlawat, S. Yadav, and A. Kumar, ''A polynomial goal programming approach for intuitionistic fuzzy portfolio optimization using entropy and higher moments,'' *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105781.

[31] M. Kara, A. Ulucan, and K. B. Atici, ''A hybrid approach for generating investor views in Black–Litterman model,'' *Expert Syst. Appl.*, vol. 128, pp. 256–270, Aug. 2019.

[32] I. B. Salehpoor and S. Molla-Alizadeh-Zavardehi, ''A constrained portfolio selection model at considering risk-adjusted measure by using hybrid meta-heuristic algorithms,'' *Appl. Soft Comput.*, vol. 75, pp. 233–253, Feb. 2019.

[33] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, ''Decision-making for financial trading: A fusion approach of machine learning and portfolio selection,'' *Expert Syst. Appl.*, vol. 115, pp. 635–655, Jan. 2019.

[34] H. H. Nesaz, M. Jasemi, and L. Monplaisir, ''A new methodology for multi-period portfolio selection based on the risk measure of lower partial moments,'' *Expert Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113032.

[35] A. H. Khan, X. Cao, V. N. Katsikis, P. Stanimirović, I. Brajević, S. Li, S. Kadry, and Y. Nam, ''Optimal portfolio management for engineering problems using nonconvex cardinality constraint: A computing perspective,'' *IEEE Access*, vol. 8, pp. 57437–57450, 2020.

[36] G. H. M. Mendonça, F. G. D. C. Ferreira, R. T. N. Cardoso, and F. V. C. Martins, ''Multi-attribute decision making applied to financial portfolio optimization problem,'' *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113527.

[37] B. Chen, J. Zhong, and Y. Chen, ''A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from Shanghai stock exchange,'' *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113104.

[38] C. B. Kalayci, O. Polat, and M. A. Akbay, ''An efficient hybrid metaheuristic algorithm for cardinality constrained portfolio optimization,'' *Swarm Evol. Comput.*, vol. 54, May 2020, Art. no. 100662.

[39] M. A. Akbay, C. B. Kalayci, and O. Polat, ''A parallel variable neighborhood search algorithm with quadratic programming for cardinality constrained portfolio optimization,'' *Knowl.-Based Syst.*, vol. 198, Jun. 2020, Art. no. 105944.

[40] W. Wang, W. Li, N. Zhang, and K. Liu, ''Portfolio formation with preselection using deep learning from long-term financial data,'' *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 113042.

[41] A. M. Aboussalah and C.-G. Lee, ''Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization,'' *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112891.

[42] M. Dhaini and N. Mansour, ''Squirrel search algorithm for portfolio optimization,'' *Expert Syst. Appl.*, vol. 178, Sep. 2021, Art. no. 114968.

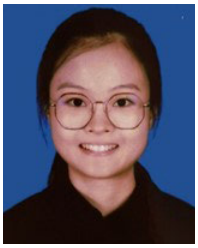
[43] Y. Dai and Z. Qin, ''Multi-period uncertain portfolio optimization model with minimum transaction lots and dynamic risk preference,'' *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107519.

[44] X. Huang, G. Jiang, P. Gupta, and M. K. Mehlawat, ''A risk index model for uncertain portfolio selection with background risk,'' *Comput. Oper. Res.*, vol. 132, Aug. 2021, Art. no. 105331.

[45] P. Gupta, M. K. Mehlawat, and A. Z. Khan, ''Multi-period portfolio optimization using coherent fuzzy numbers in a credibilistic environment,'' *Expert Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114135.

[46] J. Zhou and X. Li, ''Multi-period mean-semi-entropy portfolio management with transaction costs and bankruptcy control,'' *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 1, pp. 705–715, Jan. 2021.

[47] B. C. Barroso, R. T. N. Cardoso, and M. K. Melo, "Performance analysis of the integration between portfolio optimization and technical analysis strategies in the Brazilian stock market," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115687.
[48] X.-L. Lu and G. He, "QPSO algorithm based on Lévy flight and its application in fuzzy portfolio," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106894.
[49] T. Cura, "A rapidly converging artificial bee colony algorithm for portfolio optimization," *Knowl.-Based Syst.*, vol. 233, Dec. 2021, Art. no. 107505.
[50] T. E. Simos, S. D. Mourtas, and V. N. Katsikis, "Time-varying Black–Litterman portfolio optimization using a bio-inspired approach and neuronets," *Appl. Soft Comput.*, vol. 112, Nov. 2021, Art. no. 107767.
[51] E. K. W. Leow, B. P. Nguyen, and M. C. H. Chua, "Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation," *Expert Syst. Appl.*, vol. 179, Oct. 2021, Art. no. 115060.
[52] I. Yaman and T. E. Dalkılıç, "A hybrid approach to cardinality constraint portfolio selection problem based on nonlinear neural network and genetic algorithm," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114517.
[53] W. Chen, H. Zhang, M. K. Mehlawat, and L. Jia, "Mean–variance portfolio optimization using machine learning-based stock price prediction," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106943.
[54] M. K. Mehlawat, P. Gupta, and A. Z. Khan, "Portfolio optimization using higher moments in an uncertain random environment," *Inf. Sci.*, vol. 567, pp. 348–374, Aug. 2021.
[55] Y. Ma, R. Han, and W. Wang, "Portfolio optimization with return prediction using deep learning and machine learning," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113973.
[56] Y.-J. Liu and W.-G. Zhang, "A multi-period fuzzy portfolio optimization model with minimum transaction lots," *Eur. J. Oper. Res.*, vol. 242, no. 3, pp. 933–941, May 2015.
[57] X. Huang, "A risk index model for portfolio selection with returns subject to experts' estimations," *Fuzzy Optim. Decis. Making*, vol. 11, no. 4, pp. 451–463, Dec. 2012.
[58] M. Jain, V. Singh, and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm," *Swarm Evol. Comput.*, vol. 44, pp. 148–175, Feb. 2019.
[59] J. Sun, B. Feng, and W. Xu, "Particle swarm optimization with particles having quantum behavior," in *Proc. Congr. Evol. Comput.*, vol. 1, Jun. 2004, pp. 325–331.
[60] K. Deb, "An efficient constraint handling method for genetic algorithms," *Comput. Methods Appl. Mech. Eng.*, vol. 186, nos. 2–4, pp. 311–338, 2000.
[61] H. Zhang and G. P. Rangaiah, "An efficient constraint handling method with integrated differential evolution for numerical and engineering optimization," *Comput. Chem. Eng.*, vol. 37, pp. 74–88, Feb. 2012.
[62] Y. G. Woldesenbet, G. G. Yen, and B. G. Tessema, "Constraint handling in multiobjective evolutionary optimization," *IEEE Trans. Evol. Comput.*, vol. 13, no. 3, pp. 514–525, Mar. 2009.
[63] C. Chen and Y.-S. Zhou, "Robust multiobjective portfolio with higher moments," *Expert Syst. Appl.*, vol. 100, pp. 165–181, Jun. 2018.
[64] K. P. Anagnostopoulos and G. Mamanis, "The mean–variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14208–14217, May 2011.
[65] G.-F. Deng, W.-T. Lin, and C.-C. Lo, "Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4558–4566, Mar. 2012.
[66] K. Lwin and R. Qu, "A hybrid algorithm for constrained portfolio selection problems," *Appl. Intell.*, vol. 39, no. 2, pp. 251–266, 2013.
[67] A. Baykasoğlu, M. G. Yunusoglu, and F. B. Özsoydan, "A GRASP based solution approach to solve cardinality constrained portfolio optimization problems," *Comput. Ind. Eng.*, vol. 90, pp. 339–351, Dec. 2015.
[68] M. Mozafari, F. Jolai, and S. Tafazzoli, "A new IPSO-SA approach for cardinality constrained portfolio optimization," *Int. J. Ind. Eng. Comput.*, vol. 2, no. 2, pp. 249–262, Apr. 2011.
[69] X. Deng and X. Pan, "The research and comparison of multi-objective portfolio based on intuitionistic fuzzy optimization," *Comput. Ind. Eng.*, vol. 124, pp. 411–421, Oct. 2018.
[70] X. Li, S. Guo, and L. Yu, "Skewness of fuzzy numbers and its applications in portfolio selection," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2135–2143, Dec. 2015.
[71] T. Cura, "Particle swarm optimization approach to portfolio optimization," *Nonlinear Anal., Real World Appl.*, vol. 10, no. 4, pp. 2396–2406, Aug. 2009.
[72] C. B. Kalayci, O. Ertenlice, H. Akyer, and H. Aygoren, "An artificial bee colony algorithm with feasibility enforcement and infeasibility toleration procedures for cardinality constrained portfolio optimization," *Expert Syst. Appl.*, vol. 85, pp. 61–75, Nov. 2017.
[73] Z. Qin, "Mean-variance model for portfolio optimization problem in the simultaneous presence of random and uncertain returns," *Eur. J. Oper. Res.*, vol. 245, no. 2, pp. 480–488, Sep. 2015.
[74] J. E. Beasley, "OR-library: Distributing test problems by electronic mail," *J. Oper. Res. Soc.*, vol. 41, no. 11, pp. 1069–1072, Nov. 1990.
[75] F. Busetti, "Metaheuristic approaches to realistic portfolio optimization," 2005, *arXiv:cond-mat/0501057*.
[76] S. El-Bizri and N. Mansour, "Metaheuristics for portfolio optimization," in *Proc. Int. Conf. Swarm Intell.* Fukuoka, Japan: Springer, 2017, pp. 77–84.
[77] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. D. Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, Apr. 2003.
[78] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithm research: A history and analysis," Citeseer, Princeton, NJ, USA, Tech. Rep. TR-98-03, 1998.
[79] D. A. Van Veldhuizen, *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. Bengaluru, India: Air Force Institute of Technology, 1999.
[80] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Aug. 2002.
[81] E. Chan, *Algorithmic Trading: Winning Strategies and Their Rationale*, vol. 625. Hoboken, NJ, USA: Wiley, 2013.
[82] K. Liagkouras and K. Metaxiotis, "Multi-period mean–variance fuzzy portfolio optimization model with transaction costs," *Eng. Appl. Artif. Intell.*, vol. 67, pp. 260–269, Jan. 2018.
[83] E. Vercher, J. D. Bermúdez, and J. V. Segura, "Fuzzy portfolio optimization under downside risk measures," *Fuzzy Sets Syst.*, vol. 158, no. 7, pp. 769–782, 2007.



ZI XUAN LOKE received the B.I.T. degree (Hons.) in business computing from Universiti Malaysia Sabah, Malaysia, where she is currently pursuing the master's degree with the Faculty of Computing and Informatics. Her current research interests include meta-heuristic algorithms with a specific interest in the portfolio optimization problem.

SAY LENG GOH received the B.I.T. degree (Hons.) from Universiti Malaysia Sabah, the M.Sc. degree from Imperial College London, and the Ph.D. degree from the University of Nottingham. He is currently an Academic with the Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia. He has published in various top-tier journals, such as *European Journal of Operational Research*, *Operational Research Society*, and *Expert Systems with Applications*. His current research interests include artificial intelligence, operational research, discrete optimization, meta-heuristics, timetabling, and scheduling.

GRAHAM KENDALL is currently a Senior Vice President of the PETRA Group and the Chief Executive of the Good Capitalism Forum (GCF). Before this, he was with the University of Nottingham for 21 years. Before joining the University of Nottingham, he was in the IT sector, holding several senior positions. Before joining the PETRA Group, he was in Malaysia for ten years, being based with the University of Nottingham, Semenyih Campus, where he was the CEO and Provost and also a Pro-Vice Chancellor of Global University. He holds honorary and distinguished professorial positions at universities in Hong Kong and India. He is also an Emeritus Professor with the University of Nottingham. He has written more than 250 peer-reviewed papers. His research interests include solving real-world problems through the use of operations research, utilizing artificial intelligence methods that are based on Charles Darwin's principles of natural evolution.

He is (has been/before leaving academia) a fellow of the British Computer Society and the Operational Research Society. He is a world-recognized academic. He has served on several editorial boards, including being the Editor-in-Chief of the IEEE Transactions of Computational Intelligence and AI in Games and an Associate Editor of the *Journal of the Operational Research Society*.

SALWANI ABDULLAH received the B.Sc. degree in computer science from the University of Technology Malaysia, the master's degree in computer science from the National University of Malaysia, and the Ph.D. degree in computer science from the University of Nottingham, U.K. She is currently a Professor of computational optimization with the Faculty of Information Science and Technology, National University of Malaysia. Her research interests include artificial intelligence and operation research, particularly computational optimization algorithms (heuristics and meta-heuristics, evolutionary algorithms, and local search) that involve different real-world applications for single- and multiobjective continuous and combinatorial optimization problems, such as timetabling, scheduling, routing, nurse rostering, dynamic optimization, data mining problems (feature selection, clustering, classification, and time-series prediction), intrusion detection, and search-based software testing.

NASSER R. SABAR is currently a Lecturer with the Computer Science and IT Department, La Trobe University, Australia. He has published more than 76 papers in international journals and peer-reviewed conferences. His current research interests include hyper-heuristic frameworks, evolutionary computation, and hybrid algorithms, with a specific interest in big data optimization problems, cloud computing, dynamic optimization, and scheduling problems.