

Improving Portfolio Management Using Clustering and Particle Swarm Optimisation

Vivek Bulani¹, Marija Bezbradica¹, Martin Crane¹

¹*School of Computing, Dublin City University, Ireland*

vivek.bulani2@mail.dcu.ie, marija.bezbradica@dcu.ie, martin.crane@dcu.ie

Portfolio management, a critical application of financial market analysis, involves optimising asset allocation to maximise returns while minimising risk. This paper addresses the notable research gap in analysing historical financial data for portfolio optimisation purposes. Particularly, this research examines different approaches for handling missing values and volatility, while examining their effects on optimal portfolios. For this portfolio optimisation task, this study employs a metaheuristic approach through the Swarm Intelligence algorithm, particularly Particle Swarm Optimisation and its variants. Additionally, it aims to enhance portfolio diversity for risk minimisation by dynamically clustering and selecting appropriate assets using the proposed strategies. This entire investigation focuses on improving risk-adjusted return metrics, like Sharpe, Adjusted Sharpe, and Sortino ratios, for single-asset-class portfolios over two distinct classes of assets, cryptocurrencies and stocks. Considering relatively high market activity during pre, during and post-pandemic conditions, experiments utilise historical data spanning from 2015 to 2023. The results indicate that Sharpe ratios of portfolios across both asset classes are maximised by employing linear interpolation for missing value imputation and exponential moving average smoothing with a lower smoothing factor (α). Furthermore, incorporating assets from different clusters significantly improves risk-adjusted returns of portfolios compared to when portfolios are restricted to high market capitalisation assets.

Keywords : portfolio optimisation; clustering; asset selection; Sharpe and Adjusted Sharpe ratios; rebalancing

1 Introduction

The financial sector has been transformed by the use of algorithms to provide customised recommendations for the investment in the financial assets based on individual financial profiles and risk tolerance. This concept of finding the optimal asset combination to achieve the desired returns with risk tailored to the investors' tolerance is called *Portfolio Optimisation* (PO) [1]. A wide variety of algorithms, including the foundational Markowitz Modern Portfolio Theory (MPT) [2], Sharpe Ratio Optimisation [3], and Conditional Value at Risk (CVaR) Optimisation [4], have been suggested for this task. Recent innovations in the process of determination of appropriate weights include metaheuristic algorithms, particularly nature-inspired methods like Evolutionary and Swarm Intelligence, which evidently handle PO quite effectively under the real-world constraints [5, 6, 7].

Portfolio optimisation and management strategies are categorised as either *passive* or *active* approaches [8, 9]. Passive portfolio management involves fund managers investing client's capital in index funds or in selected securities while maintaining fixed allocations until the portfolio

reaches maturity [8]. In contrast, active portfolio management employs trading strategies to outperform benchmark market indices through strategic portfolio adjustments, such as asset selection and adjustment of weights, in response to market fluctuations [10]. Active portfolio management typically implements a rolling window methodology for the inclusion of recent market data to update the portfolio accordingly [10]. This systematic portfolio reallocation process, known as *rebalancing*, requires careful consideration of the frequency, as each rebalancing event incurs transaction costs in practical applications [11]. For both kinds of investments (active and passive), effective analysis of assets by the investor or fund manager is beneficial for the improvement of the quality of portfolios [12]. This analysis can either be classified as *fundamental analysis* or *technical analysis* [13, 14]. The former incorporates quantitative metrics and qualitative corporate indicators including product offerings, management policies, and organisational profiles [15, 16]; the latter examines inherent asset patterns through price movements, trading volumes, and market dynamics [16]. Considering the benefit of higher returns from periodically updating the portfolios rather than passive portfolio strategy [17], this research focuses exclusively on the use of technical analysis of the market for active portfolio management strategies.

The quality of market data represents a crucial component in portfolio profitability within technical market analysis frameworks. The inherent volatility of historical asset data demands processing techniques such as reduction of short term fluctuations with the help of smoothing methods like Moving Averages (MA) [18]. However, explicit research on the comparison of different preprocessing techniques for different asset types is limited. Existing studies usually employ specific methodologies for handling missing values and for smoothing purposes [19, 20], while generally restricting their analyses to shorter timeframes (i.e., few years) of historical data [21, 22, 23]. Consequently, this research focuses initially on evaluating the impact of various preprocessing methods on single-asset-type portfolios.

An important part of financial asset allocation and portfolio optimisation lies in the identification of an appropriate set of assets for the portfolio [24], which is referred to as the *asset selection stage*. This is a prerequisite, before optimising and assessing the profitability of portfolios, since the real-time financial market environment is characterised by an enormous number of financial assets of different types. Thus, to create portfolios of manageable size that consists of the most suitable assets, asset selection becomes a crucial step. Given that financial markets have been shown to exhibit cross-asset correlations [25, 26], strategic portfolio diversification has been advocated through uncorrelated asset selection as a fundamental risk mitigation approach [27]. Clustering algorithms facilitate portfolio diversification by partitioning similar assets into distinct groups [28], enabling this research to identify optimal asset selection strategies across various asset classes.

Overall, the research objective of this paper is to specifically answer the following research questions:

- RQ1: To what extent do different smoothing techniques influence risk-adjusted returns of single-asset-type portfolios of different asset classes?
- RQ2: Which selection criteria best identify representative assets from clusters formed using risk–return characteristics of smoothed data?

This research contributes by initially addressing the lack of explicit and comprehensive comparison of the effect of different missing value imputation and smoothing techniques on portfolio optimisation. Additionally, we introduce efficient asset selection strategies that effectively handle large asset space to create manageable portfolios while maintaining diversification benefits and performance characteristics.

This paper is organised as follows: Section 2 discusses the related work in the field of portfolio optimisation and how metaheuristic algorithms have been used for this purpose, with special

focus on the Particle Swarm Optimisation algorithm for optimal asset weight allocation from a large search space. Additionally, this section examines the deployment of various clustering methodologies to handle a wide variety of assets of different types. Section 3 describes the data accessing and processing steps followed by the algorithms used in our work. Further, it discusses various asset selection strategies in conjunction with the clustering algorithm and their effect on different risk-adjusted return metrics, which measure the return of an investment after consideration of the level of risk involved. A discussion on the results and other findings as well as benchmarking the results with the literature is presented in Section 4. Section 5 highlights the key outcomes and conclusions of this study, followed by potential future work mentioned in Section 6 for the creation of more realistic portfolio optimisation systems.

2 Related Work

2.1 Traditional Portfolio Optimisation Techniques

2.1.1 Markowitz Mean–Variance (MV) Theory

Markowitz’s Mean–Variance portfolio theory [2] conceptualises the trade-off between return and risk of portfolio, which are represented by their mean and variance, respectively. This enables the construction of portfolios along the *efficient frontier*, which represents the optimal portfolios that either maximise returns for given risk levels or minimise risk for given returns. Mathematically, this model can be represented as follows [5]:

$$\min \quad \sigma_{R_p}^2 = \sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{Cov}(\bar{R}_i, \bar{R}_j) \quad (1)$$

$$\text{Subject to} \quad \bar{R}_p = E(R_p) = \sum_{i=1}^N w_i \bar{R}_i \geq R \quad (2)$$

$$\sum_{i=1}^N w_i = 1 \quad (3)$$

$$w_i \geq 0, \quad \forall i \in \{1, 2, \dots, N\} \quad (4)$$

where N is the number of assets, \bar{R}_i is the mean return of asset i , $\text{Cov}(\bar{R}_i, \bar{R}_j)$ is the covariance of returns of assets i and j , and R is the investor’s target rate of return. The goal is to find the optimal weight allocation w_i that minimises the portfolio risk, σ_p^2 , for given expected return, \bar{R}_p .

2.1.2 Sharpe and Sortino Ratio

The Sharpe Ratio model [3] modifies the MV framework by incorporating risk-free returns, such as *3-Month US Treasury Bills* [29], which represents the yield from a security with theoretically zero risk. This model aims to maximise the Sharpe Ratio (SR), which is the ratio of excess returns (those exceeding the risk-free rate) to portfolio volatility (standard deviation). SR is one of the most commonly used metrics for evaluating the performance of portfolios [30, 31, 32], with higher values indicating more excess returns from the portfolio compared to the risk-free rate [21]. Mathematically, this can be represented as follows:

$$SR = \frac{\bar{r}_p - r_f}{\sigma_p} \quad (5)$$

where \bar{r}_p is the expected portfolio return, r_f is the risk-free rate of return, and σ_p is the standard deviation of the portfolio value.

The Sortino ratio (St) [33] addresses the Sharpe ratio's limitation of assuming normally distributed returns [34, 35]. St considers only the *downside deviation* [36, 37], σ_d , which is a measure used to assess investment risk by considering only the negative deviations from a target return (usually zero or the risk-free rate).

$$St = \frac{\bar{r}_p - r_f}{\sigma_d} \quad (6)$$

Another variation of Sharpe ratio, which adjusts for skewness and kurtosis in the returns distribution of the assets through the use of a penalty factor for negative skewness and excessive kurtosis, is known as the Adjusted Sharpe Ratio (ASR) [23] and is given by:

$$ASR = SR(1 + \frac{Skewness(R_p)}{6}SR - \frac{Kurtosis(R_p)}{24}SR^2) \quad (7)$$

2.2 Portfolio Optimisation Using Meta-Heuristic Algorithms

Due to the proliferation in asset data and computing power, the use of metaheuristic algorithms such as Simulated Annealing (SA) [38], Tabu search (TS) [38, 39], Particle Swarm Optimisation (PSO) [40], Ant Colony Optimisation (ACO) [41], and Evolutionary Algorithms (EA) [42] has grown. This is because research has indicated that such algorithms excel at solving diverse real-world optimisation challenges across domains such as job scheduling [43, 44], classification [45], and robotics [46]. This stems from their capability to extensively explore and exploit solution spaces while seeking globally optimal or near-optimal solutions [47]. Extensive literature reviews have been conducted on the works of portfolio management using swarm intelligence (SI) and multi-objective evolutionary algorithms during the period of 1993–2023 [48, 49, 50, 7]. They all consistently identify PSO as the predominant SI technique for Portfolio Optimisation (PO), followed by Genetic Algorithms (GAs) and Artificial Bee Colony (ABC) algorithms. To simulate the operational market conditions, previous researchers have also included different types of constraints such as bounds on holdings (limits on asset proportions), cardinality (maximum number of securities allowed), minimum transaction lots, and sector/market capitalisation constraints.

Different variations of the PSO algorithm tailored specifically for the portfolio optimisation task have been suggested by different authors [51]. A modification of PSO which increases exploration in the initial search steps and improves convergence speed in the final search steps has been developed in [52]. Using the weekly price data of different market indices, their strategy outperformed other PSO techniques in the literature on the basis of minimum mean percentage error. A novel version of PSO which considers different types of constraints discussed above is proposed in [5]. They compared their method with the Genetic Algorithm (GA) and showed that the proposed PSO effectively outperforms GA, especially in the case of large-scale problems. The criteria for evaluation used are best variance (i.e., lowest risk), mean variance, standard deviation of variances, and mean run time. In order to maintain a sufficient amount of diversity of the swarm and to prevent the PSO algorithm from premature convergence, [53] proposed a Hybrid PSO (HPSO) which maintains the diversity without computing it at every iteration, thus making the algorithm faster and efficient. A Heterogeneous Multiple Population PSO (HMPPSO) algorithm is proposed in [54], where the entire search space is partitioned into smaller sub-populations, then different variants of PSO are applied on each of these sub-population. Compared with other PSO variants in terms of mean Euclidean distance, variance of return error, and mean return error, the HMPPSO model proved to be more robust and effective, especially for high-dimensional problems. The coefficients and velocity equations have been

dynamically updated to improve the performance of PSO over other algorithms such as GA, Simulated Annealing (SA), and Tabu search (TS) [55, 38, 56]. Following the widespread use of the PSO algorithm and its variations, this work plans to use PSO and its modifications for the portfolio optimisation task.

Apart from PSO and its variants, other forms of nature-inspired algorithms have also been implemented in the past for determining optimal portfolios. A comprehensive literature review on the use of multi-objective evolutionary algorithms for the purpose of portfolio optimisation was conducted in [57], which mentions different variations of EA developed for this task. A two-stage Genetic Algorithm (GA) was introduced in [58] to initially select high-quality assets and then optimise their weights. An improved Artificial Bee Colony (ABC) algorithm which provided the right balance between exploration and exploitation to find the most optimal solution was suggested in [59]. The results obtained proved to outperform Variable Neighbourhood Search (VNS), SA, TS, and standard ABC in terms of diversity, effectiveness, and convergence. A combination of the critical components from continuous Ant Colony Optimisation (ACO), ABC, and GA models was used in [60] for solving the cardinality constrained portfolio optimisation problem. They proved their method’s efficiency over other methods including GA, TS, and SA based on percentage error and return errors. Overall, nature-inspired algorithms have been extensively used for the purpose of portfolio optimisation, accommodating various constraints inherent to financial markets. These algorithmic approaches exhibit considerable diversity in their computational complexity and implementation requirements.

2.3 Clustering of Financial Assets

Studies demonstrate that inclusion of a greater number of diverse assets in a portfolio effectively reduces unsystematic risk. This is because this kind of risk pertains specifically to individual companies or industries rather than a wider market [61]. However, increasing the count of assets has been shown to introduce the curse of dimensionality and potentially elevate transaction costs—the cost incurred while buying and/or selling financial assets. To overcome this, careful asset selection while maintaining the cardinality constraint is essential. To address this, researchers have developed preprocessing methodologies based on clustering techniques that facilitate strategic asset selection from a larger asset universe [28]. Such clustering algorithms are broadly classified into the following categories: partitional, hierarchical, density-based, and graph-based [62, 63]. Among these, k -means and k -medoids represent the most prevalent partitional algorithms, and they iteratively aim to minimise the sum of distances between observations and corresponding medoids in every cluster [64]. In contrast to k -means, where the centroids can be chosen as any arbitrary number representing the mean/average of the data objects, the k -medoids algorithm utilises actual data points/observations as centroids [65, 66], thus making it more robust to outliers and noise [67, 66]. Hierarchical clustering creates a tree-like structure by recursively merging/splitting the clusters on a distance metric. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is an implementation of a density-based algorithm that groups tightly packed financial assets according to the specified density criteria.

These different clustering algorithms and their variations have been used for the portfolio optimisation (PO) task in the literature. K -means and k -medoids algorithms followed by MV-based portfolio optimisation are used for cryptocurrency-based portfolios using the daily prices of the assets [22]. The simulation is executed for every monthly rebalancing period, wherein a single cluster, based on the risk tolerance of the investor, is selected as the optimal set of assets for the MV model. Their analysis proved that this enhances the results compared to the classic MV, risk parity, and hierarchical risk parity methods when evaluated using metrics such as Sharpe ratio, Omega ratio, Calmar ratio, Value at Risk (VaR), and Conditional Value at Risk (CVaR). VaR measures the potential loss in a portfolio over time with a specified confidence level. CVaR, also known as expected shortfall, measures tail risk by calculating the weighted average of extreme losses beyond the VaR threshold in an investment portfolio’s

return distribution [68]. Similarly, cellwise outlier robustness and agglomerative clustering has been used for cryptocurrencies [23] along with different portfolio optimisation methods such as equally weighted portfolio, minimum-variance portfolio, hierarchical risk parity [69], maxCluster portfolio, and minCluster portfolio. maxCluster [70] includes the selection of the cluster with the maximum Sharpe ratio, whereas minCluster is a novel strategy proposed by the authors where the cluster with minimum CVaR at 95% is considered. They mentioned that the minCluster portfolio has the highest Adjusted Sharpe ratio (which takes into account the skewness and kurtosis of the returns distribution) and performs 50% better than the equally weighted portfolio. In both [22, 23], portfolios are generated using the assets belonging to only a single selected cluster. *K*-means has also been used for the clustering of the top 100 cryptocurrencies in [64], followed by manual selection of 8 cryptocurrencies from these clusters for generating monthly portfolios. It considered the daily returns of the crypto coins and CVaR for finding an optimal allocation strategy. One issue that can be observed is that manual selection may not be feasible when handling large datasets and may miss some patterns or opportunities.

Apart from cryptocurrencies, clustering algorithms have been used for portfolio optimisation using data from DOW30, NASDAQ100 and S&P500 [28], the Brazilian stock exchange (B3) [66], and NSE Nifty 100 (Indian stock market) [8]. Again, *k*-means is used in [8], whereas *k*-medoids and Partitioning Around Medoids (PAM) are used in [66, 28], respectively. Similarly, *k*-means is used on financial statements of the companies to cluster the stocks [71], then assets belonging to the cluster with the highest expected return are selected for the portfolio. Then, the PSO algorithm is used for optimal weight allocation of the selected assets, while returns and Sharpe ratio metrics are used for portfolio evaluation. Hierarchical clustering with complete linkage is used to cluster industrial portfolios and S&P 500 stocks in a multi-period experimental design [72]. *K*-means, fuzzy *k*-means, and self-organising maps were used to cluster the assets of the Bombay Stock Exchange based on Euclidean distance measure, then the randomly selected assets from these clusters were used for PO using the MV model [73]. The authors concluded that *k*-means is better than other methods when evaluated based on intraclass inertia metrics like Euclidean distance and Dunn’s index. Dunn’s index evaluates the quality of a clustering solution through the compactness and separation of clusters [74]. However, the work in [73] focuses on single-period portfolios, whereas in real time, portfolios are updated after a certain period of time to increase profitability [23].

All of these works have concluded that, irrespective of the assets and performance metrics used, clustering does help to generate better portfolios than other traditional methods based on mean variance model. This also aligns with the concept of modern portfolio theory, which suggests that well-diversified asset universe can further minimise risk and improve MV model’s efficiency. To overcome the shortcomings mentioned above in the previous works and improve them to make more realistic portfolio optimisation (PO) framework, this work aims to implement an automated multi-period PO method which selects assets from different clusters, as done by many of the previous researchers [75, 76, 64], without any manual intervention.

3 Data Handling and Asset Selection Strategies

The study begins with identification of the most effective preprocessing methodology for historical data of cryptocurrencies and stocks. This involves investigating the impact of various imputing methods for missing values and moving average techniques for data smoothing on corresponding single-asset-class portfolios. The analysis of data preprocessing techniques is conducted on a constrained set of top market capitalisation assets (as in [21, 23]) from the two asset classes, in order to generate optimal portfolios while maintaining reasonable transaction costs and portfolio cardinality [24, 23]. Subsequent analysis involves clustering larger sets of available assets and identifying a suitable selection strategy for diverse portfolio generation. This entire work uses the Particle Swarm Optimisation (PSO) algorithm for finding the optimal weight allocation of the

assets in the portfolio. Evaluation of portfolio performance is performed using the risk-adjusted return metrics such as Sharpe, Sortino, and Adjusted Sharpe ratios given in Equations (5)–(7). The values of these metrics have been annualised for easier interpretation. The downside deviation (σ_d) for the Sortino ratio is calculated using the formula proposed in [77].

3.1 Dataset Description

The dataset [78] utilised for this work is a historic dataset spanning from January 2005 to May 2023. It contains transaction data for selected assets from several categories: US Stocks, ETFs, Futures, US Indices, and Cryptocurrencies. The dataset presents these data at various granularities/frequencies in the OHLCV format, i.e., Date/Time, Open Price, High Price, Low Price, Close Price, Volume of shares. For this work, the focus is on hourly data ranging from January 2018 to May 2023 for Cryptocurrencies, and the daily data from January 2015 to May 2023 for the S&P500 and Nasdaq100 stocks. These particular timeframes have been selected due to considerably high market activity, observed in the plots of assets’ historical data, during this time period compared to most of the preceding periods, while also encompassing both pre-pandemic and post-pandemic market trends. The use of S&P500 and Nasdaq100 stocks in this study is inspired from previous works such as [1, 28, 79]. The risk-free rate of return considered for Sharpe, Adjusted Sharpe, and Sortino ratios is the commonly used 3-Month US Treasury Bill Secondary Market Rate, Discount Basis (TB3MS) data, which is freely accessible online at [80]. This downloaded data includes the percent, non-adjusted returns of monthly frequency.

The initial analysis task employs a small subset from the large pool of available assets, comprising the top 10 cryptocurrency coins by market capitalisation according to <https://coinmarketcap.com> (accessed on 9th April 2024), and the top 20 S&P stocks by market capitalisation index weight according to [81] (accessed on 24th March 2024), all listed in Table 3 (Appendix A). Subsequently, for the analysis of clustering and asset selection methodologies, the scope expands to include the complete sets of Nasdaq100, S&P 500, and all 53 available cryptocurrencies in the dataset.

3.2 Dataset Preprocessing

As the risk-free data is composed of 3-month treasury returns, hence in order to accurately calculate risk-adjusted returns, cumulative 3-month returns must be computed for each asset to determine excess returns relative to this risk-free rate. Following general practice [82], returns from OHLC (Open, High, Low, Close) data are derived using assets’ closing prices by calculating the percentage changes between two time points [83], as shown in Equation (8). Hence, in order to find cumulative quarterly returns, first monthly returns are calculated and then aggregated over the duration of three months.

$$\text{Daily Return} = \frac{Price_{(today)} - Price_{(yesterday)}}{Price_{(yesterday)}} \quad (8)$$

3.2.1 Handling Missing Values

Whenever there is a zero volume, i.e., no transaction/trading has occurred, such entries are not recorded in the dataset. Hence in this work, in order to deal with such gaps, global replacement of the missing values with the corresponding minimum, maximum, mean (average), etc., of the time series [84, 85] has been compared to interpolation methods. The interpolation approach in this study employs linear and quadratic regression methods to model these trends and predict suitable values for missing values imputation. For the comparison of these different methods, initially, $p\%$ (=“percent_missing”) of the available data is replaced with NaN values (as per [86]), followed by prediction of these missing values using various imputation techniques. The

predicted values are then compared with the true values for evaluation purposes. This evaluation is performed using error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), as also performed by [86, 87, 88]. Lower values of these metrics indicate superior method performance. In this investigation, the different “percent_missing” ($p\%$) values used are [10%, 30%, 50%, 70%, 90%] [88, 86].

3.2.2 Implementation of Smoothing Algorithms

To overcome the issue of high volatility in asset prices, smoothing of data is performed after handling the missing values. In general, the literature lacks a comprehensive analysis of how different smoothing strategies impact portfolio performance [89], as measured by risk-adjusted return metrics like the Sharpe ratio. Therefore, as part of RQ1, this study examines these effects across single-asset-type portfolios from various asset classes. The analysis considers several smoothing techniques, including Simple Moving Average (SMA), Exponential Moving Average (EMA), 4-point Moving Average (FMA), and 2×4 -point Moving Average (TFMA) [13, 90, 91]. They are formulated as follows :

$$SMA_n(t) = \frac{\sum_{i=1}^n X(t-i+1)}{n} \quad (9)$$

$$EMA_n(t) = \alpha \cdot X(t) + (1 - \alpha) \cdot EMA_n(t-1) \quad \text{given that} \quad EMA_n(1) = X(1) \quad (10)$$

$$FMA(t) = \frac{X(t-2) + X(t-1) + X(t) + X(t+1)}{4} \quad (11)$$

$$TFMA(t) = \frac{FMA(t) + FMA(t+1)}{2} \quad (12)$$

where (n) is the window size, $\alpha \in [0, 1]$ is the *smoothing factor*, and is related to n by $\alpha = 2/(n+1)$. While SMA assigns equal weights to all the past n data points, EMA assigns exponentially decreasing weights to previous data points. A larger α emphasises recent observations, whereas a smaller α value assigns greater weight to historical data points as well. This work implements and compares multiple window sizes for SMA ($n = [5, 10, 30, 50, 100]$) and different smoothing factors for EMA ($\alpha = [0.01, 0.1, 0.3, 0.5, 0.9]$).

3.3 Meta-Heuristic Algorithm Used for Portfolio Optimisation—Particle Swarm Optimisation

Particle Swarm Optimisation (PSO) is characterised by its simplicity—attributed to a reduced number of parameters—and its rapid convergence even in the presence of large search spaces [48]. Hence, this research employs the PSO methodology to determine optimal asset allocation weights within the portfolio framework. The Particle Swarm Optimisation (PSO) [40] algorithm was developed by Kennedy and Eberhart in 1995, which was inspired by the movements of swarms in nature such as flock of birds and schools of fish. The PSO algorithm identifies the optimal solution by comprehensively exploring the objective function space through systematic adjustment of the routes of individual agents, called *particles*. Each particle in the swarm has inherent properties such as position and velocity, which help them to move towards an optimal position. In the context of portfolio optimisation, the particle’s position in the swarm represents the weight allocations (i.e., portfolio), and its velocity represents the rate of change in the individual weights [92]. Every particle i has a tendency to move randomly to search through an unexplored search space. But at the same time, each particle is attracted towards its personal best position P_i until the present iteration as well as the current global best position P_g of the swarm. In other words,

P_i indicates a particle's position in the entire history for which the objective function value is maximum or minimum (depending on the problem requirement). The aim of the PSO algorithm is to discover the optimal position or solution in a search space, considering the best solutions associated with each individual particle. This process continues until either the objective function value ceases to improve or a predefined iteration limit is reached.

Mathematically, the equations to update the velocity $V_i^t = (v_{i,1}^t, v_{i,2}^t, \dots, v_{i,D}^t)^\top$ and position $X_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,D}^t)^\top$ of the particle i at iteration t , using the personal best position $P_i^t = (p_{i,1}^t, p_{i,2}^t, \dots, p_{i,D}^t)^\top$ and global best position $P_g^t = (p_{g,1}^t, p_{g,2}^t, \dots, p_{g,D}^t)^\top$, are given as follows [93]:

$$v_{i,d}^{t+1} = \omega v_{i,d}^t + c_1 r_{1,i,d}^t (p_{i,d}^t - x_{i,d}^t) + c_2 r_{2,i,d}^t (p_{g,d}^t - x_{i,d}^t) \quad (13)$$

$$x_{i,d}^{t+1} = x_{i,d}^t + v_{i,d}^{t+1} \quad (14)$$

where

$d = 1, 2, \dots, D$ represent dimensions; $i = 1, 2, \dots, N$ represent the particle;

N is the size of the swarm, i.e., the total number of particles; ω is the inertia weight;

c_1, c_2 are two positive constants, called the cognitive and social parameters, respectively;

$r_{1,i,d}, r_{2,i,d}$ are random numbers, uniformly distributed in $[0, 1]$;

g is the index of the overall best particle in the swarm; and

$t = 1, 2, \dots$ determines the iteration number of the algorithm.

The velocity calculation in Equation (13) comprises three components [93]: the inertia component ($\omega v_{i,d}^t$), which maintains directional momentum; the cognitive component ($c_1 r_{1,i,d}^t (p_{i,d}^t - x_{i,d}^t)$), which attracts particles to their historical best positions; and the social component ($c_2 r_{2,i,d}^t (p_{g,d}^t - x_{i,d}^t)$), which pulls particles towards the neighbourhood's global best position. Figure 1 represents the particle's movement from position X_i^t to X_i^{t+1} , where P_i^t represents particle i 's best position at step t and $P_g^t \approx \min(P_i^t) (\forall i = 1, 2, \dots, N)$ denotes the current global best at step t . In the case of search space problems, there are the two important concepts of *exploration* and *exploitation*. Exploration investigates unexplored regions for new possibilities, while exploitation focuses on previously visited points' neighbourhoods to maximise immediate gains [94]. Excessive exploration increases computational time, whereas pure exploitation yields sub-optimal global solutions, necessitating a balance to efficiently identify optimal solutions while avoiding local optima. In Equation (13), parameters c_1 and c_2 govern PSO's exploration-exploitation balance: higher cognitive component values ($c_1 > c_2$) promote wider particle wandering, while higher social component values ($c_1 < c_2$) increase premature convergence risk to local minima [40]. Research advocates a dynamic approach where exploration is prioritised during initial search phases, while inter-particle communication intensifies as the algorithm approaches termination [93].

The implementation in this work encompasses three variations of the PSO algorithm, each employing different mechanisms for updating particle position and velocity. They are Standard PSO (SPSO) [40], Improved PSO (IPSO) [93], and Drift PSO (DPSO) [32]. These variations correspond to different ways of searching through the search space to identify the most optimal asset weight allocations, which are represented by the particle with best fitness value. A fitness function, also called an objective function, is used to evaluate the quality of each particle. In the application of portfolio optimisation, the risk-adjusted return metrics are used as the fitness function.

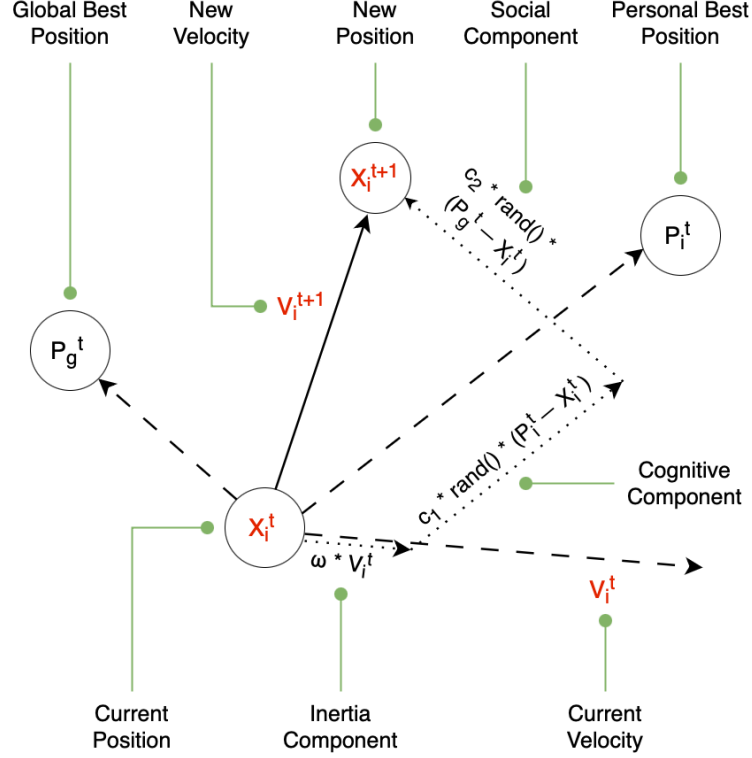


Figure 1: Adaptation of [13] with detailed explanation of the components to show the update of a particle's position in PSO. Here, * represents the standard multiplication of the terms.

The pseudo code for the standard PSO algorithm is shown in Algorithm A1 (Appendix B).

One of the drawbacks of the Standard PSO algorithm is that it has a high tendency to become trapped in the local optimum due to its rapid convergence property [95]. To overcome this, works such as [93] dynamically adjust the hyper-parameters such as inertia weights (ω) and constants (c_1, c_2) in IPSO using Equations (15)–(17) instead of keeping them fixed as in SPSO. As discussed before, higher c_1 values correspond to more exploration, and higher c_2 values result in more exploitation. Therefore, in (16) and (17), c_1 decreases and c_2 increases as iterations progress, encouraging early exploration followed by later-stage exploitation [93]. This can potentially improve the efficiency of finding optimal asset weight allocations.

$$\omega = 0.81 - \frac{t}{t_{max}} \cdot 0.4 \quad (15)$$

$$c_1 = 1.0 - \frac{t}{t_{max}} \quad (16)$$

$$c_2 = 1.0 + \frac{t}{t_{max}} \quad (17)$$

where t is the number of iterations and t_{max} is the maximum number of iterations.

Additionally, at every iteration of IPSO, the particle exhibiting the minimum fitness value (risk-adjusted return metric) is redirected to move in the opposite direction of its velocity, as proposed by [93]. These particles' velocity and position updates follow Equations (18) and (19), while the remaining particles use standard Equations (13) and (14).

$$v_{i,d}^{t+1} = \omega v_{i,d}^t + c_1 r_{1,i,d}^t (p_{i,d}^t - x_{i,d}^t) + c_2 r_{2,i,d}^t (p_{g,d}^t - x_{i,d}^t) \quad (18)$$

$$x_{i,d}^{t+1} = x_{i,d}^t - v_{i,d}^{t+1} \quad (19)$$

Lastly, the DPSO algorithm used in this study follows the works in [32], where the algorithm draws inspiration from electron motion in conductors under electric fields, incorporating both thermo motion and drift motion. Thermo motion refers to the random movement of electrons due to thermal energy, whereas drift motion refers to electrons' directional movement under external electric fields. In [32], thermo motion is considered to be following Maxwell's velocity distribution law. Thus, it is represented as $\sigma_{i,d}^t \cdot \psi_{i,d}^t$ where $\sigma_{i,d}^t$ is the standard deviation of the Gaussian distribution and $\psi_{i,d}^t$ is a random number with the standard normal distribution. The drift motion is represented as $c_1 r_{1,i,d}^t (p_{i,d}^t - x_{i,d}^t) + c_2 r_{2,i,d}^t (p_{g,d}^t - x_{i,d}^t)$ for particle i at iteration $t + 1$. Consequently, the final equation for updating the velocity and position of particle i is as follows:

$$v_{i,d}^{t+1} = \alpha |L_d^t - x_{i,d}^t| \psi_{i,d}^t + c_1 r_{1,i,d}^t (p_{i,d}^t - x_{i,d}^t) + c_2 r_{2,i,d}^t (p_{g,d}^t - x_{i,d}^t) \quad (20)$$

$$x_{i,d}^{t+1} = x_{i,d}^t + v_{i,d}^{t+1} \quad (21)$$

where

$d = 1, 2, \dots, D$ represent dimensions;

$i = 1, 2, \dots, N$ represent particles;

N is the size of the swarm, i.e., total number of particles;

α is called compression–expansion coefficient;

$\psi_{i,d}$ is a random number from a standard normal distribution $N(0, 1)$;

and

$L^t = (L_1^t, L_2^t, \dots, L_D^t)$ is mean best (*mbest*), i.e., average of *pbest* of all particles at iteration t ,

i.e., $L_d^t = (1/N) \sum_{i=1}^N p_{i,d}^t (\forall d = 1, 2, \dots, D)$

All these variants of PSO have two critical hyperparameters—the number of iterations and the population size. Population size indicates the number of particles that make up the set of the entire search space, and each of these particles' position represents a portfolio vector. A comprehensive analysis of the impact of different values for these hyperparameters on corresponding generated portfolios is performed using a range of values of 200–500 for the number of iterations and 100–400 for the population size.

To validate the effectiveness of these PSO algorithms, the results are benchmarked against previous works. In the literature, PSO has been compared with the genetic algorithm based on the Sharpe ratio of the portfolios constructed from Shanghai Stock Exchange 50 Index constituents [30]. In addition, the Sharpe ratio has been used to optimise financial portfolios using a proposed variation of PSO in [31]. Following the use of the Sharpe ratio as a metric in such papers, this research initially compares the PSO algorithm with Markowitz portfolio theory based on the optimal Sharpe ratio for cryptocurrency-only portfolios, with no rebalancing scenario. This is performed using an inbuilt library in Python called PyPortfolioOpt [96], which implements the Markowitz mean–variance model. This is done with the help of the `max_sharpe()` function, which belongs to this library's class called `EfficientFrontier` [97, 98, 99]. Further analysis demonstrates the superiority of these PSO methods over some of the other existing approaches in both cryptocurrency and stock-based portfolio optimisation.

3.4 K-Medoids-Based Clustering and Optimal Selection of Financial Assets

As a part of RQ2, this study aims to handle larger sets of cryptocurrencies and stocks in real time using clustering algorithms followed by identification of the most suitable assets to be included in the portfolio. A distance-based clustering approach is selected for its simplicity, intuitive nature, and broad applicability [100]. The process aims to identify the most profitable and

diverse assets to maximise returns while minimising risk, as illustrated in Figure 2. Based on the considerations and literature works reviewed previously, the K -medoids algorithm is adopted as the clustering technique. For this clustering task, assets are represented using a bivariate approach incorporating quarterly returns and standard deviation of returns, consistent with the methodology employed in [100].

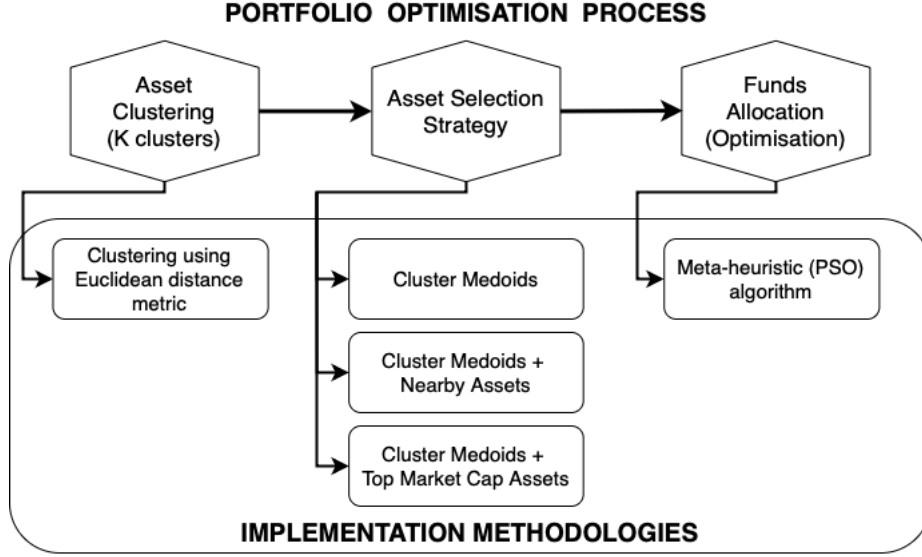


Figure 2: Process representing the three steps of PO that include clustering and asset selection.

Partitional clustering algorithms require prior knowledge of the number of clusters (K) [100], which presents challenges, particularly with large datasets [64]. According to [101], the optimal value of K should be determined using domain expertise and/or statistical methods like the Silhouette plot. In view of this, this research applies common approaches for identifying the optimal K value: the Elbow method, which minimises the total within-cluster sum of squares (WSS) [102], alongside Silhouette analysis [103]. Once the optimal number of clusters has been decided, the K -medoids algorithm [65] is executed, which starts with the initialisation of the K medoids. Two types of initialisation methods for the centroids have been tested in this work: *random*, where centroids are randomly selected [28, 64], and *heuristic*, where points with the smallest sum of distance to every other point are chosen [104]. The pseudo code for the K -medoids clustering algorithm used is shown in Algorithm A2 (Appendix B).

Once the clustering of assets has been performed, this research proposes following novel strategies for asset selection for the portfolio. These methods differ in terms of how additional assets are selected to complement the representative medoid assets and the degree of flexibility permitted during this selection process.

(a) Asset Selection Strategy 1: The methodology exclusively selects the centroids of the clusters created using Euclidean distance as a metric. The underlying principle posits that these centroids serve as optimal representatives of the assets within each cluster, consequently leading to effective portfolio construction. Hence, through this strategy, the total number of assets = number of clusters (K).

(b) Asset Selection Strategy 2: Research indicates that portfolios comprising approximately 10–15 assets typically yield optimal results [64, 24, 8]. This methodology therefore incorporates, in addition to centroids, additional P nearest assets to each centroid from every cluster such that the size of the portfolios created is approximately 10–15. Here, the total number of selected assets equals the number of clusters (K) + P .

(c) Asset Selection Strategy 3: This approach represents an enhancement of Strategy 2 by incorporating a performance-driven, dynamic asset selection mechanism. Rather than selecting

a predetermined number of assets from each cluster, the method automatically identifies the best-performing candidates based on multiple evaluation criteria. Given the required final portfolio size S , the model identifies S nearest assets to centroids from each cluster and ranks them individually based on four metrics: returns, risk, risk-adjusted returns, and distances from corresponding centroids. Assets are sorted in decreasing order of returns and risk-adjusted returns, and increasing order of risk and distance to centroids. The frequency with which each asset appears in the top S positions of these individual ranking lists is calculated. Lastly, the top S assets with the highest cumulative ranking frequency are selected as the final set of assets. This methodology enables dynamic selection of the most suitable assets from different clusters based on multiple performance indicators.

(d) Asset Selection Strategy 4: In this strategy, the idea is that there are usually some top market capitalisation-based assets across different asset classes. Investment in such assets can prove advantageous for particular investor profiles due to their characteristic stability [105]. Based on this principle, this strategy incorporates the top Q performing assets (ranked by market capitalisation) during the considered time period, in addition to the medoid assets. Consequently, the total number of assets selected for portfolio optimisation = number of clusters (K) + Q .

In order to make the portfolios more realistic and dynamic, the portfolios usually need to be revised after certain amounts of time, and this is termed *rebalancing*. It enables portfolios to remain aligned with the market trends and potentially generate higher profits compared to the *Buy and Hold* strategy [106, 107] (which involves a single investment maintained over an extended period). Rebalancing facilitates monitoring of portfolio value fluctuations over time. However, it incurs transaction costs while buying or selling securities in the financial markets, due to which determination of optimal rebalancing frequency becomes critical. This research establishes a fixed rebalancing interval of $l = 30$ days as per [23, 108, 22]. This approach exemplifies *static rebalancing*, wherein adjustments occur after predetermined time intervals. During each rebalancing period, the optimal value of K is first determined, followed by application of clustering and asset selection techniques to adjust portfolios according to market trends.

4 Results

4.1 Missing Value Handling Techniques

The comparison of missing value handling methods utilises complete datasets of Nasdaq100, S&P 500 indices, and all 53 cryptocurrencies. Considering these asset types, different percentages of missing values ($p\%$), and different imputation strategies discussed before, linear interpolation (as also used by [19]) demonstrates superior performance, exhibiting minimal values for Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) metrics. Consequently, linear interpolation is selected as the optimal technique for addressing missing values in both cryptocurrency and stock datasets throughout the subsequent analyses.

4.2 Analysis of Different Smoothing Strategies

To examine the effects of various smoothing techniques on portfolio quality, a comparative analysis of Sharpe ratios was conducted using different window sizes (n) for Simple Moving Average (SMA) and varying smoothing factor (α) for Exponential Moving Average (EMA) along with other smoothing techniques such as 4-point moving average (FMA) and 2×4 -point moving average (TFMA). As discussed earlier, this investigation utilises the top 10 cryptocurrencies and top 20 S&P stocks based on market capitalisation, which are detailed in Table 3 (Appendix A). No portfolio rebalancing occurs throughout this analytical process.

The results for SMA with different window sizes (n) reveal that the best Sharpe ratio values are obtained using $n = 50$ for cryptocurrency and $n = 199$ for S&P stocks. On the other hand, EMA

provides the best Sharpe ratio values with smoothing factor $\alpha = 0.1$ for cryptocurrency and $\alpha = 0.01$ for S&P stocks, both corresponding to larger window sizes. Overall, the comparison of the different smoothing methods mentioned above illustrates that Exponential Moving Average (EMA) emerges as the optimal smoothing technique for both asset classes (cryptocurrencies and stocks) during the timeframe of data considered. However, the optimal smoothing factor varies between asset types, with cryptocurrency portfolios performing best at $\alpha = 0.1$ (which gives the optimal Sharpe ratio of around 0.76) and S&P stock-based portfolios at $\alpha = 0.01$ (which gives the optimal Sharpe ratio of around 2.43). The lower value of α (which indicates greater window size) for S&P stocks than for cryptocurrencies can be attributed to the higher volatility of cryptocurrencies as compared to stocks [109, 110, 111]. These findings on the optimal smoothing factor values align with existing literature (like [112]), which recommends that medium to long-term day traders, in general, should employ lower α values. For day traders, the smoothing factor considered for EMA can be increased. However, this research focuses on medium to long-term trading strategies using extended historical data. Hence, EMA ($\alpha = 0.1$) for cryptocurrency data and EMA ($\alpha = 0.01$) for stock data are implemented throughout subsequent analyses to effectively eliminate random market fluctuations.

4.3 Hyperparameter Optimisation for the Particle Swarm Optimisation (PSO) Algorithm

In order to determine the most efficient configuration for the three PSO variants used in this research, multiple values for the “iterations” and “population size” hyperparameters are systematically evaluated. Due to the reasons mentioned previously, this analysis considers the top 10 cryptocurrency coins and top 20 S&P stocks based on market capitalisation (Table 3). There is no rebalancing scenario considered for this analysis; hence, the complete data for these assets for the entire considered time period are used at once. Based on the Sharpe ratios of portfolios generated from various PSO configurations, a consistent pattern is observed wherein portfolio performance for both SPSO and IPSO models demonstrates a positive correlation with increases in population size and iteration count. However, in order to balance algorithmic efficiency against rate of improvement in portfolio quality, 500 iterations and a population size of 400 are established as the most efficient hyperparameter values for the upcoming analyses.

4.4 Benchmarking PSO with Previous Works

In this section, the effectiveness of PSO is compared against other established techniques in portfolio optimisation. Initially, the PSO algorithm is compared with Markowitz theory using Python’s inbuilt library called Efficient Frontier. The assets considered comprise the top 10 cryptocurrencies based on market capitalisation (Table 3), with their complete data being used at once, as there is no rebalancing. As the average risk-free rate of the 3-month US Treasury bill during January 2018 to May 2023 is approximately 1.5, the risk-free rate parameter is set to 1.5 for both the `max_sharpe()` function of the Efficient Frontier library and risk-adjusted return calculations in PSO techniques. Under these conditions, the Efficient Frontier Library is unable to generate any portfolio and states that “at least one of the assets must have an expected return exceeding the risk-free rate” [113]. But, PSO algorithms can produce optimal portfolios with Sharpe ratios of approximately 0.56. Furthermore, PSO algorithms are also observed to consistently outperform the Efficient Frontier library across various other risk-free rate scenarios, both higher and lower than the 1.5% rate.

The PSO methods are subsequently benchmarked against the methods used in other works focusing on both cryptocurrency and stock-based portfolios [21, 114]. Initial comparative analysis focuses on cryptocurrency portfolio optimisation, utilising the work of [114] as a benchmark. Different models used in this paper are the Equally Weighted (NAIVE) portfolio, Mean Variance model (MV), maximum Sharpe model (Max Sharp), mean conditional value-at-risk (MCVaR)-

based models, and the multicriteria (MC)-based model. There are two MV-based models developed in [114]—MV Middle, which focuses on the average variance, and MV Max, which aims to obtain the maximum variance. Their MC model is a multiobjective decision-making model that considers a number of variables such as daily return, standard deviation, volume, market capitalisation, etc. These models are used to create an optimal portfolio from the top nine cryptocurrencies as per market capitalisation (based on rankings during the entire timeframe of data)—Bitcoin (BTC), Dash, Ethereum Classic (ETC), Ethereum (ETH), Litecoin (LTC), Monero (XMR), Neo, Stellar (XLM) and Ripple (XRP). The timeframe considered in their paper spans from 1 January 2017 to 11 February 2020, with monthly rebalancing resulting in a total of 32 rebalancing time periods/portfolios. The metrics used for PSO training are annualised Sharpe, Adjusted Sharpe, and Sortino ratios, with the risk-free rate set to 0 as per [114]. Three evaluation metrics from the paper are used for the comparative analysis : return for the next training day (next-day returns), mean return for the next 30 days (mean 30 returns), and standard deviation of the next 30 days returns (SD 30 returns). These results are presented in figs. 3 to 5, which illustrate the number of rebalancing periods for which our PSO variations performed better than the models employed in [114]. The larger the number of these rebalancing periods, the better the performance of our PSO algorithms compared to the benchmark methods. Analysis of these graphs indicates that when evaluated based on returns-based metrics (i.e., next-day returns and mean 30 returns), the PSO methods outperform all benchmark models for approximately more than half of the 32 rebalancing time periods. When assessed based on risk (i.e., SD 30 returns), the PSO methods demonstrate superior performance compared to all benchmark models across all 32 periods.

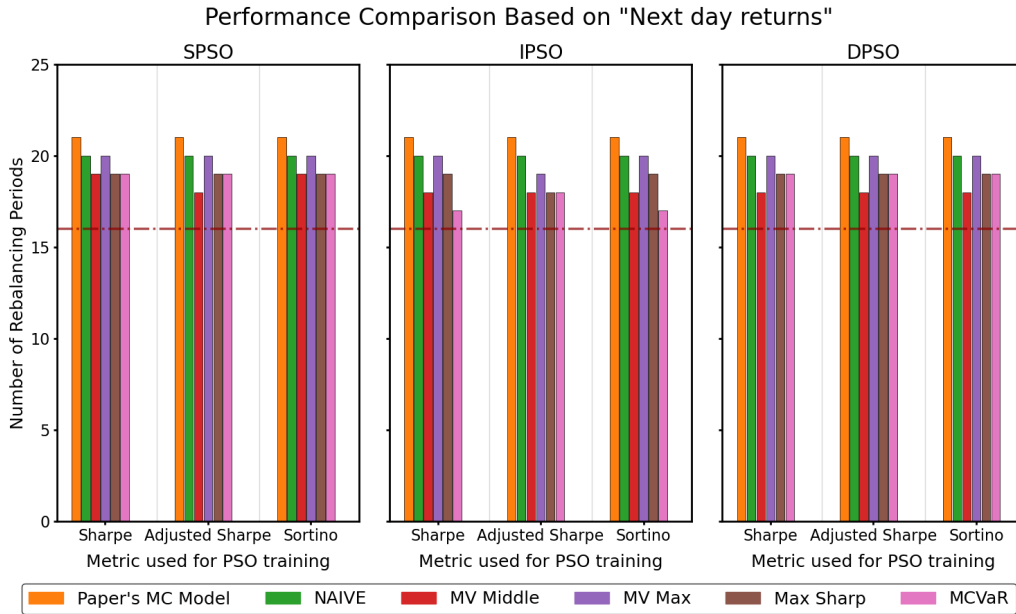


Figure 3: Number of time periods between 1 January 2017 and 11 February 2020 for which the PSO models in this research were better than the 6 models mentioned in the paper based on the next-day returns evaluation metric. In total, rebalancing took place over 32 time periods. Different risk-adjusted return metrics were used for training the PSO models.

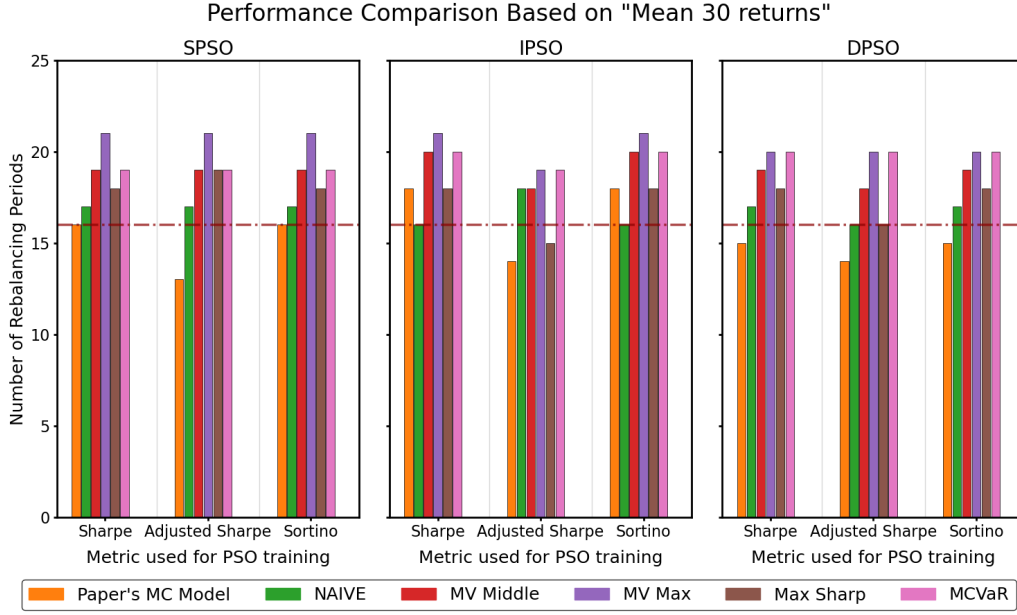


Figure 4: Number of time periods between 1 January 2017 and 11 February 2020 for which the PSO models in this research were better than the 6 models mentioned in the paper based on the mean 30 returns evaluation metric. In total, rebalancing took place over 32 time periods. Different risk-adjusted return metrics were used for training the PSO models.

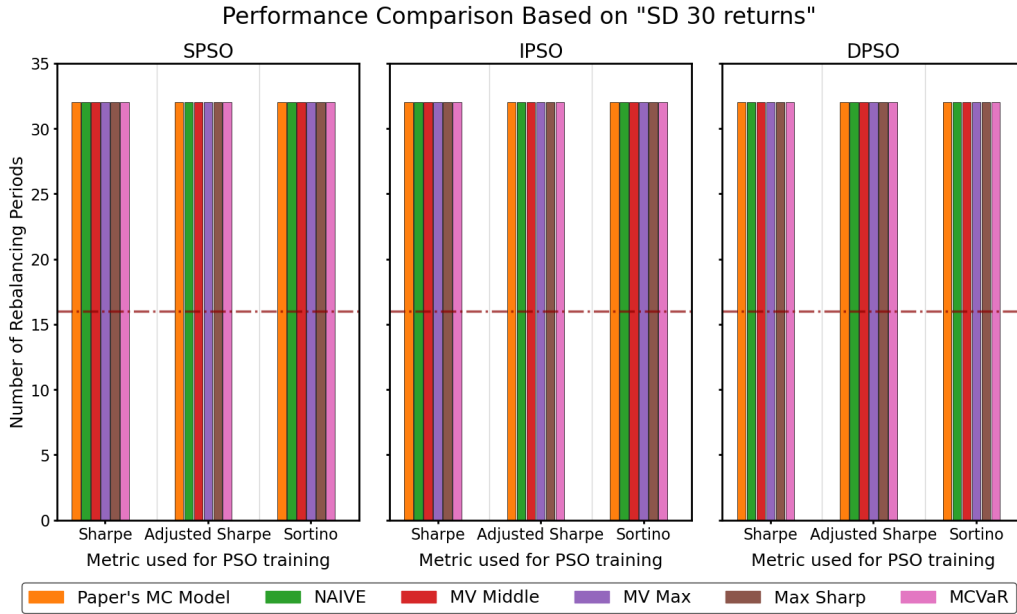


Figure 5: Number of time periods between 1 January 2017 and 11 February 2020 for which the PSO models in this research were better than the 6 models mentioned in the paper based on the SD 30 returns evaluation metric. In total, rebalancing took place over 32 time periods. Different risk-adjusted return metrics were used for training the PSO models.

Lastly, the evaluation of the PSO variations for stock-based portfolios involves benchmarking against [21]. The results are compared with those presented in Table 6 of the paper (where no selling of assets is permitted). These results consider the following stocks for portfolio optimisation: Berkshire Hathaway Inc (BRKa), JPMorgan Chase & co. (JPM), Johnson & Johnson (JNJ), Procter & Gamble Company (PG), Visa Inc Class A (V). The comparison is

presented in Table 1, which demonstrates that the PSO techniques combined with the exponential moving average smoothing strategy (with $\alpha = 0.01$) gives significantly better Sharpe ratio values for the stocks-only portfolio.

Table 1: Comparison of Sharpe ratio values for stocks-only portfolios obtained using PSO techniques and smoothing methods in this research work against the benchmark paper [21].

Stocks Only			
SPSO	IPSO	DPSO	Paper
4.8832	4.8802	4.8843	1.27

4.5 Analysis of the Effects of Clustering and Different Asset Selection Techniques

To view the effects of clustering of the assets followed by the selection of optimal assets using the four novel strategies proposed above, clustering and portfolio optimisation is performed for every rebalancing period to create dynamic portfolios. This task utilises complete datasets of Nasdaq100 and all 53 available cryptocurrencies. Nasdaq100 stocks are used instead of top 20 S&P stocks to increase the size of the asset pool and demonstrate the effectiveness of the clustering and asset selection methods. The training period encompasses January 2018 to December 2022 for cryptocurrencies and January 2015 to December 2022 for stocks, constituting approximately 80–85% of the total data. Then monthly portfolio rebalancing commences in January 2022 and continues until May 2023. Asset selection strategies 2, 3, and 4 are implemented in a way that they generate portfolios containing approximately 12–15 assets, thus facilitating simple and efficient portfolio management while maintaining reasonable transaction costs [24, 23]. For asset selection strategy 4, which requires the inclusion of top-performing assets based on market capitalisation, coinmarketcap.com [115] serves as the reference for an updated list of top cryptocurrencies during each rebalancing time period. Similarly, the top 10 Nasdaq100 stocks are identified using market capitalisation rankings [116, 115].

An assessment of elbow graphs and silhouette scores across every rebalancing period for both asset types reveals the trends depicted in Figures 6 and 7. Hence, across every rebalancing period, a cluster count of $k = 4$ for cryptocurrencies and $k = 3$ for Nasdaq100 stocks is implemented consistently for all clustering methodologies. Cluster analysis at every rebalancing period indicates a distinct separation between high-return/high-risk asset clusters and low-return/low-risk asset clusters, validating that the selection of assets from each cluster effectively enhances portfolio diversity.

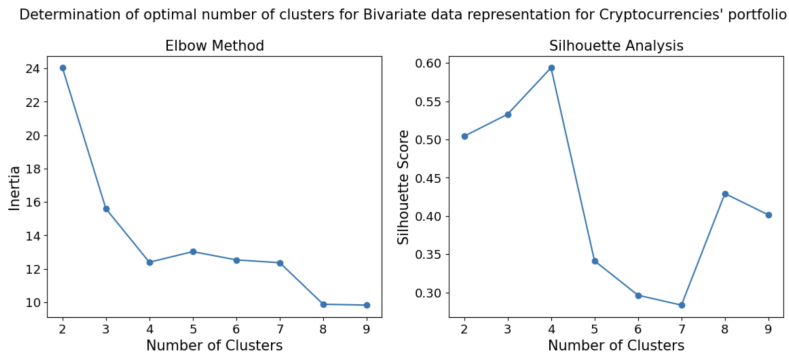


Figure 6: Elbow graph and silhouette scores for different values of K (number of clusters) for cryptocurrency assets.

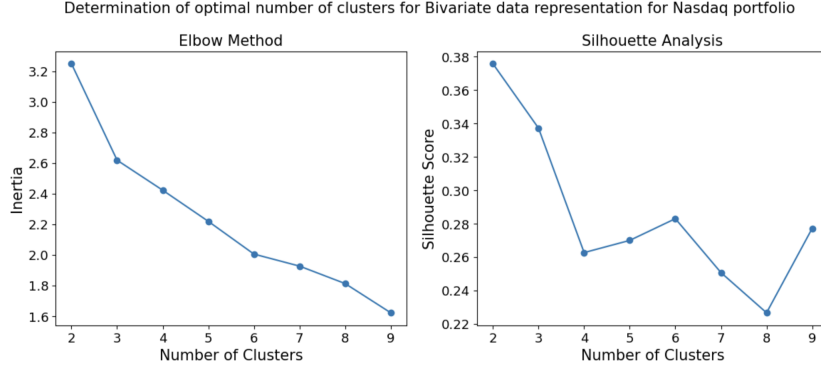


Figure 7: Elbow graph and silhouette scores for different values of K (number of clusters) for Nasdaq100 stocks.

An evaluation of portfolios constructed using assets from different selection strategies (such as only the centroids, or centroids + nearby assets, or centroids + top market capitalisation assets), was conducted based on Sharpe, Sortino, and Adjusted Sharpe ratios. The results for the Sharpe and Adjusted Sharpe ratios for the cryptocurrency-based portfolios are shown in Figure 8, and the results for the Nasdaq100-based portfolios are shown in Figure 9 (similar trends have been found with the Sortino ratio as well). These figures provide graphical representations of descriptive statistics of risk-adjusted returns across the four asset selection variations for cryptocurrency-only and stock-only portfolios. The statistics include minimum, maximum, and average (i.e., mean) of the metrics calculated across all monthly rebalancing periods. Optimal performance is characterised by higher minimum, maximum, and average metric values.

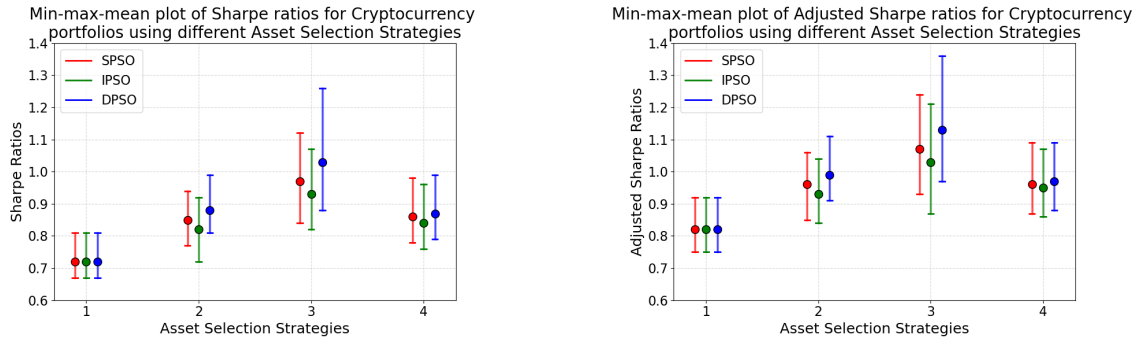


Figure 8: Min-max-mean plot of risk-adjusted return metrics for cryptocurrency portfolios using the 4 strategies proposed for clustering + asset selection. The left and right subplots illustrate the minimum, maximum, and mean values of the Sharpe and Adjusted Sharpe ratios, respectively, while considering portfolios during all rebalancing periods.

4.5.1 Comparison of the Effects of Clustering and Asset Selection Strategy with the Non-Clustered Approach on the Corresponding Portfolios

Figure 8 shows that all clustering and asset selection techniques generally yield improved Sharpe ratios compared to when portfolio optimisation is applied to smoothed data without performing any clustering (Sharpe ratio of 0.76 is obtained for top market capitalisation cryptocurrency-based portfolios without clustering while analysing different smoothing strategies in the previous section). Also this comparison illustrates that instead of creating portfolios from just the high market capitalisation assets (as was done in the previous section of analysis of smoothing strategies), performing clustering and including other diverse set of assets from these clusters helps to improve the performance of the portfolios. These findings can also be observed from

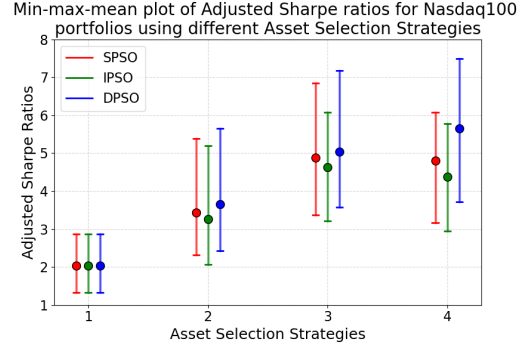
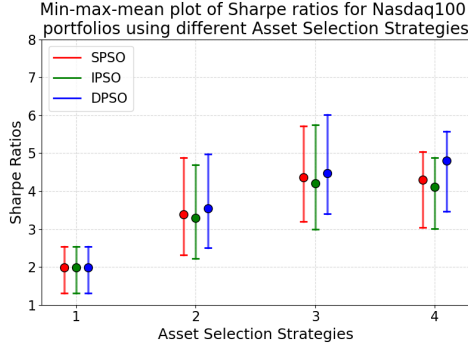


Figure 9: Min-max-mean plot of risk-adjusted return metrics for Nasdaq100 portfolios using the 4 strategies proposed for clustering + asset selection. The left and right subplots illustrate the minimum, maximum, and mean values of the Sharpe and Adjusted Sharpe ratios, respectively, while considering portfolios during all rebalancing periods.

Figure 8 and they align with the initial proposed hypothesis. Similar conclusions can be formed on the basis of the Adjusted Sharpe ratio from Figure 8.

4.5.2 Comparison of Different PSO Techniques

Figures 8 and 9 show that among the three PSO variations, Drift PSO (DPSO) consistently generates the most optimal portfolios, irrespective of the risk-adjusted return metric used. This is also observed in the results obtained during the implementation of different smoothing strategies. Figures 10 and 11 also provide graphical confirmation of this conclusion, demonstrating variations in risk-adjusted return metrics over time across different PSO algorithms for cryptocurrency-based and stock-based portfolios, respectively.

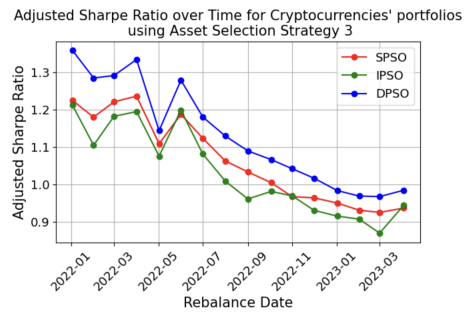
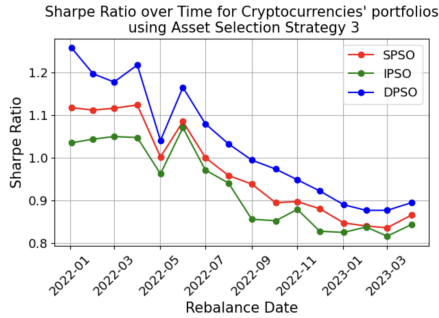


Figure 10: Comparison of PSO algorithms for cryptocurrency-based portfolios obtained using asset selection method 3, optimised using the Sharpe and Adjusted Sharpe ratios.

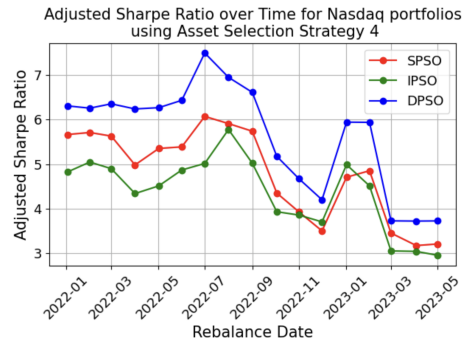
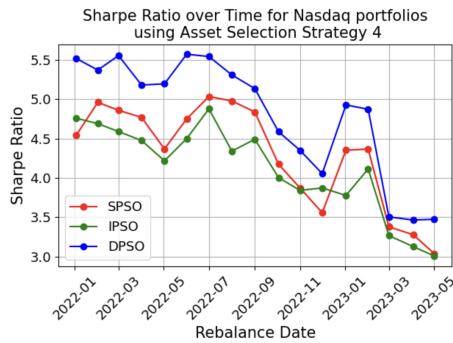


Figure 11: Comparison of PSO algorithms for Nasdaq100-based portfolios obtained using asset selection method 4 optimised using the Sharpe and Adjusted Sharpe ratios.

4.5.3 Comparison of Different Asset Selection Strategies

Figure 8 clearly indicates that strategy 3 provides the best Sharpe and Adjusted Sharpe ratio values for cryptocurrency portfolios (Similar results were found for the Sortino ratio as well). This demonstrates that enabling the algorithm to dynamically select cryptocurrencies from different clusters based on different performance metrics (including return, risk, return/risk, distance to centroid) significantly enhances portfolio risk-adjusted returns. This can also be seen in Figures 12 and 13, where the graphs represent the variation in Sharpe and Adjusted Sharpe ratios of the cryptocurrency portfolios over time obtained from the DPSO algorithm for different asset selection methods. Only the results from the DPSO method are depicted in these graphs because, as shown before, this algorithm is found to consistently perform the best amongst the three PSO variations.

Conversely, in Figure 9, for Nasdaq100 stock portfolios, both asset selection strategy 4 (incorporating top market capitalisation stocks alongside centroids) and strategy 3 (dynamic asset selection) generate comparably optimal portfolios. This pattern is also graphically illustrated in Figures 14 and 15, which display trends in the Sharpe and Adjusted Sharpe ratios of the Nasdaq100 portfolios over time from the DPSO algorithm for different asset selection methods.

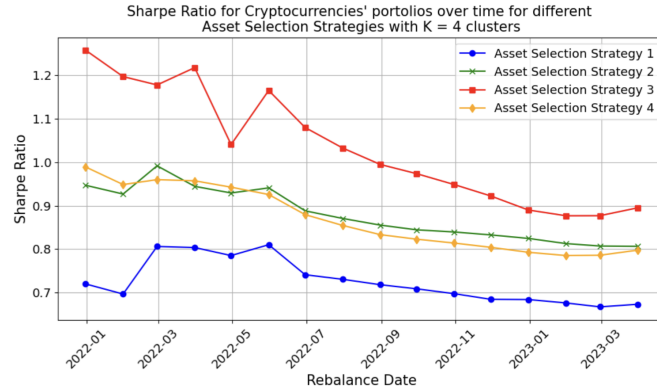


Figure 12: Comparison of different asset selection strategies for cryptocurrency-based portfolios obtained using the DPSO algorithm using the Sharpe ratio.

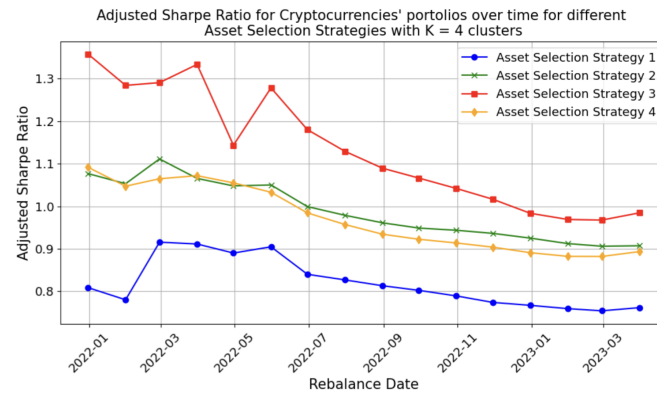


Figure 13: Comparison of different asset selection strategies for cryptocurrency-based portfolios obtained using the DPSO algorithm using the Adjusted Sharpe ratio.

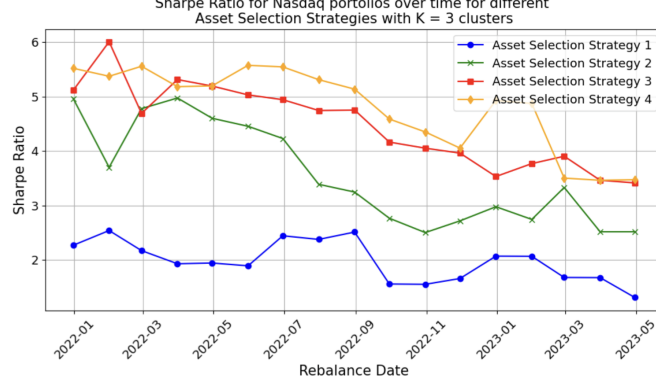


Figure 14: Comparison of different asset selection strategies for Nasdaq100-based portfolios obtained using the DPSO algorithm using the Sharpe ratio.

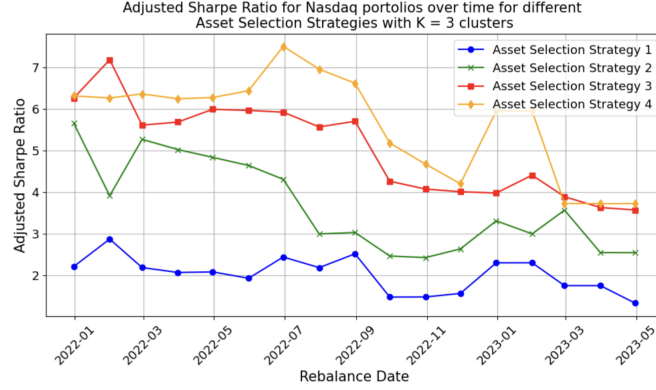


Figure 15: Comparison of different asset selection strategies for Nasdaq100-based portfolios obtained using the DPSO algorithm using the Adjusted Sharpe ratio.

4.6 Benchmarking with Literature Review

To benchmark the clustering results, the methodology's performance is compared against [77]. The comparison examines the Sharpe and Sortino ratio values for the portfolios of sizes $(n) = 10$ and $(n) = 25$ in Table 2. This comparison is conducted against PSO techniques of the restricted Sharpe and restricted Sortino models for equivalent portfolio sizes in the benchmark paper. The dataset comprises daily prices of S&P 500 stocks from October 2017 to October 2018. For the clustering approaches in this work, $k = n$ is considered for asset selection method 1, whereas for the remaining strategies, $k = 5$ for $n = 10$ and $k = 10$ for $n = 25$ are considered based on the elbow graphs and silhouette scores. Table 2 presents the results from the DPSO algorithm, which typically yields optimal portfolios. Evidence indicates that all the proposed asset selection strategies in this work outperformed those in the reference paper. Additionally, for both cases of portfolio size, dynamic selection of assets from clusters demonstrated superior performance.

Table 2: Comparison of Sharpe and Sortino ratios for different asset selection strategies with the benchmark paper for 2 kinds of S&P 500-based portfolios, where the number of assets (n) = 10 or 25. Values below the diagonal correspond to Sharpe ratios, and values above the diagonal correspond to Sortino ratios. Purple and green highlighted cells represent the best Sharpe and Sortino ratio values respectively for different portfolio sizes. The PSO algorithm used here is DPSO.

	Sortino	Asset Select 1	Asset Select 2	Asset Select 3	Asset Select 4	Paper
Sharpe						
$n = 10$		3.31	3.22	3.47	3.24	1.81
	12.51		12.44	13.12	12.36	1.837
$n = 25$		3.50	3.44	3.79	3.10	2.398
	12.84		13.03	14.83	11.73	2.717

5 Discussion and Conclusions

Portfolio optimisation has evolved through machine learning, deep learning, and more recently, swarm intelligence and evolutionary algorithms for efficient asset management. Particle Swarm Optimisation (PSO) is a nature-inspired approach that demonstrates proficiency with continuous data. It offers simplicity through fewer parameters, and rapid convergence in large search spaces helps it to effectively address financial market challenges. This study implemented three different variations of the PSO algorithm to enhance the efficacy of optimal solution identification within the search space. This work initially addressed the management of missing data and high volatility in the historical data of financial assets. For this, comparison of different imputation and smoothing strategies was implemented for different types of assets, particularly cryptocurrencies and US stocks. It was found that linear interpolation for missing value imputation followed by the exponential moving average smoothing technique with a smaller smoothing factor (α) value helped the most to improve the quality of cryptocurrency-based and stocks-based portfolios, measured in terms of the Sharpe ratio. Also, the smoothing factor for cryptocurrency was higher than for the stocks data, which corresponded to a smaller window size for cryptocurrency than for stocks for medium to long-term investments. This is due to higher volatility of the considered cryptocurrencies compared to stocks. Apart from this, the superiority of the PSO algorithm for both asset type portfolios over the Markowitz mean-variance model and other traditional algorithms used in the literature was also demonstrated. Subsequent analysis in this work focused on clustering of assets using a partition-based clustering algorithm and identifying the most suitable asset selection strategy from different proposed strategies for the two asset classes. The overall results demonstrated that dynamic selection of assets from different clusters based on multiple performance metrics, such as return, risk, risk-adjusted returns, and distance to the centroids, helps to improve the Sharpe and Adjusted Sharpe ratios of both cryptocurrency and stocks-based portfolios. This indicates that restricting the selection of assets based on a single metric, such as only including high market capitalisation assets, diminishes portfolio quality. Increasing the diversity of portfolios through careful inclusion of high-return/high-risk and low-return/low-risk assets maximises the profitability of the constructed portfolios.

6 Future Work

The main focus of this work was on analysing and using historical market data for generating optimal portfolios with the help of artificial intelligence methods. However there is a significant impact of social media and influencers on the prices of assets, especially in the case of cryptocurrency. The relationship between transactions from some top-rated cryptocurrencies

such as Bitcoin, Ethereum, and Ripple with the activity in online forums was analysed in [117]. The results showed that the number of cryptocurrency transactions is greatly influenced by the comments and replies posted in forums and online communities. Thus, considering the influence of social media activity when generating optimal portfolios is important.

Though PSO has shown improvements in performance with the help of parallel computations, it will likely become trapped in local optima due to premature convergence in the case of complex problems. This issue may be encountered when the search space is more scattered. However, this limitation of PSO may be overcome by using it along with other evolutionary approaches such as genetic algorithms.

Author Contributions : Conceptualization, V.B., M.B., and M.C.; Methodology, V.B., M.B., and M.C.; Software, V.B.; Formal Analysis, V.B.; Investigation, V.B.; Data Curation, V.B.; Writing—Original Draft Preparation, V.B.; Writing—Review and Editing, M.B. and M.C.; Visualization, V.B.; Supervision, M.B. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding : This publication has emanated from research conducted with the financial support of Taighde Éireann-Research Ireland under Grant No. 18/CRT/6223.

Data availability : The data can be shared upon reasonable request.

Acknowledgments : Vivek Bulani wishes to acknowledge the financial support of Taighde Éireann-Research Ireland under Grant No. 18/CRT/6223. (URL: <https://www.crt-ai.ie/> (accessed on 28th March 2025)). Martin Crane and Marija Bezbradica wish to acknowledge partial support from the ADAPT, the Research Ireland Centre for AI-Driven Digital Content Technology at DCU [13/RC/2106_P2] (URL: <https://www.adaptcentre.ie/> (accessed on 28th March 2025)). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest : The authors declare no conflicts of interest.

A Subset of Assets Used

Table 3: List of subset of assets considered.

Top 10 Crypto Coins	Top 20 S&P Stocks	Top 20 S&P Stocks
Bitcoin (BTC)	MICROSOFT CORP (MSFT)	APPLE INC (AAPL)
Ethereum (ETH)	NVIDIA CORP (NVDA)	AMAZON.COM, INC (AMZN)
Tether (USDT)	META PLATFORMS INC, CLASS A (META)	ALPHABET INC CL C (GOOG)
Ripple (XRP)	BERKSHIRE HATHAWAY INC. CL B (BRK.B)	ELI LILLY AND COMPANY (LLY)
USD Coin (USDC)	BROADCOM INC. (AVGO)	TESLA, INC (TSLA)
Dogecoin (DOGE)	JPMORGAN CHASE & COMPANY (JPM)	UNITEDHEALTH GROUP INC (UNH)
Cardano (ADA)	VISA INC. (V)	EXXON MOBIL CORP (XOM)
Tron (TRX)	JOHNSON & JOHNSON (JNJ)	MASTERCARD INC (MA)
Litecoin (LTC)	THE PROCTER & GAMBLE COMPANY (PG)	HOME DEPOT, INC. (HD)
Dai (DAI)	MERCK COMPANY. INC. (MRK)	COSTCO WHOLESALE CORP (COST)

B Pseudocodes

B.1

Algorithm A1: Standard Particle Swarm Optimisation (SPSO) Algorithm

Data: Population of particles with positions and velocities, fitness function

Result: Optimal solution represented by particle with the maximum fitness value

Initialise :

Maximum number of iterations (max_it), population size (pop_size), Inertia weight (ω), cognitive parameter (c_1) and social parameter (c_2)

foreach *particle* **do**

Randomly initialise the position with values within the predefined range;
Randomly initialise the velocity with values within the predefined range;
Calculate the objective function value (also called as fitness value) for each particle. In terms of portfolio optimisation, this represents metrics like Sharpe, Adjusted Sharpe or Sortino ratios;
Set personal best position to the current position;
Set personal best fitness to the current fitness;

Set global best position to the best personal best position;

Set global best fitness to the best personal best fitness;

Optimisation Loop :

while *termination criteria not met* **do**

foreach *particle* (i) **do**

Set *timeout* = 10

while $x_i(t+1)$ is not within the predefined range AND *timeout* > 0 **do**

Randomly initialise r_{1_i} and r_{2_i} from $U(0,1)$

Update the Velocity :

$$v_i(t+1) = \omega \cdot v_i(t) + c_1 \cdot r_{1_i}(t) \cdot (pbest_i(t) - x_i(t)) + c_2 \cdot r_{2_i}(t) \cdot (gbest(t) - x_i(t));$$

Update the Position :

$$x_i(t+1) = x_i(t) + v_i(t+1);$$

timeout = *timeout* - 1

Evaluate the Fitness :

Evaluate fitness of the new position of the particle;

Update the Personal Best :

if *new fitness better than personal best* **then**

Update the personal best position;

Update the personal best fitness;

Update the Global Best :

foreach *particle* (i) **do**

if *fitness better than global best* **then**

Update the global best position;

Update the global best fitness;

Return:

Return global best position, i.e., the particle which has the highest fitness value (Sharpe, Adjusted Sharpe or Sortino ratio). This particle represents the optimal portfolio as the optimised solution;

B.2

Algorithm A2: K-Medoids Clustering Algorithm

Data: Dataset D with n objects, number of clusters K , maximum number of iterations

Result: K clusters and their medoids

Initialise :

 Select K objects from D as the initial medoids, either randomly or heuristically;

Iterative Clustering :

while *medoids change OR maximum number of iterations not reached* **do**

Assign / Update Clusters :

foreach *object* $o_i \in D$ **do**

 Assign o_i to the nearest medoid;

Update Medoids :

foreach *cluster* C_i **do**

foreach *object* $o_j \in C_i$ **do**

 Calculate the total cost of o_j as the medoid. This means to calculate the total distance of o_j to every other object in the cluster C_i ;

 Select the object o_j with the minimal total cost as the new medoid for C_i ;

 Calculate the total cost for the cluster (i.e., within cluster sum of squares) ;

Return:

 Return K clusters and their medoids;

References

- [1] Ta, V.D.; Liu, C.M.; Tadesse, D.A. Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Appl. Sci.* **2020**, *10*, 437. <https://doi.org/10.3390/app10020437>.
- [2] Markowitz, H.M.; *Portfolio Selection: Efficient Diversification of Investments*; J. Wiley: Hoboken, NJ, USA, 1959.
- [3] Sharpe, W.F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *J. Financ.* **1964**, *19*, 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>.
- [4] Rockafellar, R.T.; Uryasev, S. Optimization of conditional value-at-risk. *J. Risk* **2000**, *2*, 21–41. <https://doi.org/10.21314/JOR.2000.038>.
- [5] Golmakani, H.R.; Fazel, M. Constrained Portfolio Selection using Particle Swarm Optimization. *Expert Syst. Appl.* **2011**, *38*, 8327–8335. <https://doi.org/10.1016/j.eswa.2011.01.020>.
- [6] Niu, B.; Fan, Y.; Xiao, H.; Xue, B. Bacterial foraging based approaches to portfolio optimization with liquidity risk. *Neurocomputing* **2012**, *98*, 90–100. <https://doi.org/10.1016/j.neucom.2011.05.048>.
- [7] Metaxiotis, K.; Liagkouras, K. Multiobjective Evolutionary Algorithms for Portfolio Management: A comprehensive literature review. *Expert Syst. Appl.* **2012**, *39*, 11685–11698. <https://doi.org/10.1016/j.eswa.2012.04.053>.

- [8] Aithal, P.K.; Geetha, M.; U, D.; Savitha, B.; Menon, P. Real-Time Portfolio Management System Utilizing Machine Learning Techniques. *IEEE Access* **2023**, *11*, 32545–32559. <https://doi.org/10.1109/ACCESS.2023.3263260>.
- [9] Gunjan, A.; Bhattacharyya, S. A brief review of portfolio optimization techniques. *Artif. Intell. Rev.* **2023**, *56*, 3847–3886. <https://doi.org/10.1007/s10462-022-10273-7>.
- [10] Grinold, R.C.; Kahn, R.N. *Active Portfolio Management*; McGraw-Hill: New York, 2000.
- [11] El Bernoussi, R.; Rockinger, M. Rebalancing with transaction costs: Theory, simulations, and actual data. *Financ. Mark. Portf. Manag.* **2023**, *37*, 121–160.
- [12] S, K. *Security Analysis and Portfolio Management*, 3rd ed.; PHI Learning Pvt. Ltd.: Delhi, India, 2022.
- [13] Thakkar, A.; Chaudhari, K. A Comprehensive Survey on Portfolio Optimization, Stock Price and Trend Prediction Using Particle Swarm Optimization. *Arch. Comput. Methods Eng.* **2021**, *28*, 2133–2164. <https://doi.org/10.1007/s11831-020-09448-8>.
- [14] Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A systematic review of fundamental and technical analysis of stock market predictions. *Artif. Intell. Rev.* **2020**, *53*, 3007–3057. <https://doi.org/10.1007/s10462-019-09754-z>.
- [15] Thakkar, A.; Chaudhari, K. CREST: Cross-Reference to Exchange-based Stock Trend Prediction using Long Short-Term Memory. *Procedia Comput. Sci.* **2020**, *167*, 616–625. <https://doi.org/10.1016/j.procs.2020.03.328>.
- [16] Anbalagan, T.; Maheswari, S.U. Classification and Prediction of Stock Market Index Based on Fuzzy Metagraph. *Procedia Comput. Sci.* **2015**, *47*, 214–221. <https://doi.org/10.1016/j.procs.2015.03.200>.
- [17] Wen, Q.; Yang, Z.; Song, Y.; Jia, P. Automatic stock decision support system based on box theory and SVM algorithm. *Expert Syst. Appl.* **2010**, *37*, 1015–1022.
- [18] Raudys, A.; Lenčiauskas, V.; Malčius, E. Moving averages for financial data smoothing. In Proceedings of the Information and Software Technologies: 19th International Conference, ICIST 2013, Kaunas, Lithuania, 10–11 October 2013; Proceedings 19; Springer: Berlin/Heidelberg, Germany, 2013; pp. 34–45.
- [19] Cesarone, F.; Scozzari, A.; Tardella, F. A new method for mean-variance portfolio optimization with cardinality constraints. *Ann. Oper. Res.* **2013**, *205*, 213–234.
- [20] Lim, Q.Y.E.; Cao, Q.; Quek, C. Dynamic portfolio rebalancing through reinforcement learning. *Neural Comput. Appl.* **2022**, *34*, 7125–7139.
- [21] Ma, Y.; Ahmad, F.; Liu, M.; Wang, Z. Portfolio optimization in the era of digital financialization using cryptocurrencies. *Technol. Forecast. Soc. Change* **2020**, *161*, 120265. <https://doi.org/10.1016/j.techfore.2020.120265>.
- [22] Lorenzo, L.; Arroyo, J. Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm. *Financ. Innov.* **2023**, *9*, 25. <https://doi.org/10.1186/s40854-022-00438-2>.
- [23] Menvouta, E.J.; Serneels, S.; Verdonck, T. Portfolio optimization using cellwise robust association measures and clustering methods with application to highly volatile markets. *J. Financ. Data Sci.* **2023**, *9*, 100097. <https://doi.org/10.1016/j.jfds.2023.100097>.

- [24] Maghsoodi, A.I. Cryptocurrency portfolio allocation using a novel hybrid and predictive big data decision support system. *Omega* **2023**, *115*, 102787. <https://doi.org/10.1016/j.omega.2022.102787>.
- [25] McMillan, D.G. Cross-asset relations, correlations and economic implications. *Glob. Financ. J.* **2019**, *41*, 60–78.
- [26] Zeevi, A.; Mashal, R. Beyond Correlation: Extreme Co-Movements Between Financial Assets. Available at SSRN 317122. 2002. Available online: <https://ssrn.com/abstract=317122> (accessed on 20th March 2025).
- [27] Koumou, G.B. Diversification and portfolio theory: A review. *Financ. Mark. Portf. Manag.* **2020**, *34*, 267–312.
- [28] Tolun Tayalı, S. A novel backtesting methodology for clustering in mean–variance portfolio optimization. *Knowl.-Based Syst.* **2020**, *209*, 106454. <https://doi.org/10.1016/j.knsys.2020.106454>.
- [29] U.S. Department of the Treasury. Available online: https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_bill_rates&field_tdr_date_value=2023. (accessed on 9th April 2024).
- [30] Zhu, H.; Wang, Y.; Wang, K.; Chen, Y. Particle Swarm Optimization (PSO) for the constrained portfolio optimization problem. *Expert Syst. Appl.* **2011**, *38*, 10161–10169.
- [31] Zaheer, K.B.; Abd Aziz, M.I.B.; Kashif, A.N.; Raza, S.M.M. Two stage portfolio selection and optimization model with the hybrid particle swarm optimization. *Matematika* **2018**, *34*, 125–141.
- [32] Sun, J.; Fang, W.; Wu, X.; Lai, C.H.; Xu, W. Solving the multi-stage portfolio optimization problem with a novel particle swarm optimization. *Expert Syst. Appl.* **2011**, *38*, 6727–6735. <https://doi.org/10.1016/j.eswa.2010.11.061>.
- [33] Sortino, F.A.; Price, L.N. Performance measurement in a downside risk framework. *J. Invest.* **1994**, *3*, 59–64.
- [34] Bailey, D.H.; Lopez de Prado, M. The Sharpe ratio efficient frontier. *J. Risk* **2012**, *15*, 13.
- [35] Mistry, J.; Shah, J. Dealing with the limitations of the Sharpe ratio for portfolio evaluation. *J. Commer. Account. Res.* **2013**, *2*, 10.
- [36] Cuchieri, N. Deep Reinforcement Learning for Financial Portfolio Optimisation. Master’s Thesis, University of Malta, Msida, Malta, 2021.
- [37] Sharma, A.; Mehra, A. Financial analysis based sectoral portfolio optimization under second order stochastic dominance. *Ann. Oper. Res.* **2017**, *256*, 171–197.
- [38] Chang, T.J.; Meade, N.; Beasley, J.E.; Sharaiha, Y.M. Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.* **2000**, *27*, 1271–1302. [https://doi.org/10.1016/S0305-0548\(99\)00074-X](https://doi.org/10.1016/S0305-0548(99)00074-X).
- [39] Schaerf, A. Local Search Techniques for Constrained Portfolio Selection Problems. *Comput. Econ.* **2002**, *20*, 177–190. <https://doi.org/10.1023/A:1020920706534>.
- [40] Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN’95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>.

- [41] Dorigo, M.; Maniezzo, V.; Colorni, A. Ant system: Optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man, Cybern. Part B (Cybern.)* **1996**, *26*, 29–41. <https://doi.org/10.1109/3477.484436>.
- [42] Vikhar, P.A. Evolutionary algorithms: A critical review and its future prospects. In Proceedings of the 2016 International Conference on Global Trends in Signal Processing, information computing and communication (ICGTSPICC), Jalgaon, India, 22–24 December 2016; pp. 261–265.
- [43] Lin, T.L.; Horng, S.J.; Kao, T.W.; Chen, Y.H.; Run, R.S.; Chen, R.J.; Lai, J.L.; Kuo, I.H. An efficient job-shop scheduling algorithm based on particle swarm optimization. *Expert Syst. Appl.* **2010**, *37*, 2629–2636. <https://doi.org/10.1016/j.eswa.2009.08.015>.
- [44] Nguyen, S.; Zhang, M.; Johnston, M.; Tan, K.C. Automatic Programming via Iterated Local Search for Dynamic Job Shop Scheduling. *IEEE Trans. Cybern.* **2015**, *45*, 1–14. <https://doi.org/10.1109/TCYB.2014.2317488>.
- [45] Chernbumroong, S.; Cang, S.; Yu, H. Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 282–289. <https://doi.org/10.1109/JBHI.2014.2313473>.
- [46] Chen, C.H.; Liu, T.K.; Chou, J.H. A Novel Crowding Genetic Algorithm and Its Applications to Manufacturing Robots. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1705–1716. <https://doi.org/10.1109/TII.2014.2316638>.
- [47] Yang, X.S.; Talatahari, S.; Alavi, A.H. *Metaheuristic Applications in Structures and Infrastructures*; Elsevier : Waltham, MA, USA, 2013.
- [48] Ertenlice, O.; Kalayci, C.B. A survey of swarm intelligence for portfolio optimization: Algorithms and applications. *Swarm Evol. Comput.* **2018**, *39*, 36–52. <https://doi.org/10.1016/j.swevo.2018.01.009>.
- [49] Chen, Y.; Zhao, X.; Yuan, J. Swarm intelligence algorithms for portfolio optimization problems: Overview and recent advances. *Mob. Inf. Syst.* **2022**, *2022*, 4241049.
- [50] Erwin, K.; Engelbrecht, A. Meta-heuristics for portfolio optimization. *Soft Comput.* **2023**, *27*, 19045–19073.
- [51] Leung, M.F.; Wang, J. Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 2825–2836.
- [52] Deng, G.F.; Lin, W.T.; Lo, C.C. Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization. *Expert Syst. Appl.* **2012**, *39*, 4558–4566. <https://doi.org/10.1016/j.eswa.2011.09.129>.
- [53] Wang, W.; Wang, H.; Wu, Z.; Dai, H. A Simple and Fast Particle Swarm Optimization and Its Application on Portfolio Selection. In Proceedings of the 2009 International Workshop on Intelligent Systems and Applications, Wuhan, China, 23–24 May 2009; pp. 1–4. <https://doi.org/10.1109/IWISA.2009.5072675>.
- [54] Yin, X.; Ni, Q.; Zhai, Y. A novel PSO for portfolio optimization based on heterogeneous multiple population strategy. In Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 25–28 May 2015; pp. 1196–1203. <https://doi.org/10.1109/CEC.2015.7257025>.

- [55] Chen, C.; Chen, B.y. Complex portfolio selection using improving particle swarm optimization approach. In Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 828–835.
- [56] Koshino, M.; Murata, H.; Kimura, H. Improved particle swarm optimization and application to portfolio selection. *Electron. Commun. Jpn. (Part III Fundam. Electron. Sci.)* **2007**, *90*, 13–25.
- [57] Ponsich, A.; Jaimes, A.L.; Coello, C.A.C. A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. *IEEE Trans. Evol. Comput.* **2012**, *17*, 321–344.
- [58] Yu, L.; Wang, S.; Lai, K.K. Portfolio optimization using evolutionary algorithms. In *Reflexing Interfaces: The Complex Coevolution of Information Technology Ecosystems*; IGI Global: Hershey, PA, USA, 2008; pp. 235–245.
- [59] Chen, A.H.L.; Liang, Y.C.; Liu, C.C. Portfolio optimization using improved artificial bee colony approach. In Proceedings of the 2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), Singapore, 16–19 April 2013; pp. 60–67. <https://doi.org/10.1109/CIFEr.2013.6611698>.
- [60] Kalayci, C.B.; Polat, O.; Akbay, M.A. An efficient hybrid metaheuristic algorithm for cardinality constrained portfolio optimization. *Swarm Evol. Comput.* **2020**, *54*, 100662.
- [61] Machdar, N.M. The Effect of Capital Structure, Systematic Risk, and Unsystematic Risk on Stock Return. *Bus. Entrep. Rev.* **2015**, *14*, 149–160.
- [62] Rodriguez, M.Z.; Comin, C.H.; Casanova, D.; Bruno, O.M.; Amancio, D.R.; Costa, L.d.F.; Rodrigues, F.A. Clustering algorithms: A comparative approach. *PLoS ONE* **2019**, *14*, e0210236.
- [63] Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160.
- [64] Tenkam, H.M.; Mba, J.C.; Mwambi, S.M. Optimization and Diversification of Cryptocurrency Portfolios: A Composite Copula-Based Approach. *Appl. Sci.* **2022**, *12*, 6408. <https://doi.org/10.3390/app12136408>.
- [65] Rduseeun, L.; Kaufman, P. Clustering by means of medoids. In Proceedings of the Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, Switzerland, 31 August–4 September 1987; Volume 31.
- [66] Duarte, F.G.; De Castro, L.N. A Framework to Perform Asset Allocation Based on Partitional Clustering. *IEEE Access* **2020**, *8*, 110775–110788. <https://doi.org/10.1109/ACCESS.2020.3001944>.
- [67] Arora, P.; Deepali.; Varshney, S. Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia Comput. Sci.* **2016**, *78*, 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>.
- [68] Cui, X.; Sun, X.; Zhu, S.; Jiang, R.; Li, D. Portfolio optimization with nonparametric value at risk: A block coordinate descent method. *INFORMS J. Comput.* **2018**, *30*, 454–471.
- [69] Lopez de Prado, M. Building diversified portfolios that outperform out-of-sample. *J. Portf. Manag.* **2016**.

- [70] Sass, J.; Thös, A.K. Risk reduction and portfolio optimization using clustering methods. *Econom. Stat.* **2024**, *32*, 1–16.
- [71] U, I.; Yun, I.; Jong, H.; Rim, W. Portfolio Optimization Based on K-Means Clustering and Particle Swarm Optimization Using Financial Statements and Stock Price Data. Available online: <https://ssrn.com/abstract=4937613> (accessed on 28th March 2025).
- [72] Bjerring, T.T.; Ross, O.; Weissensteiner, A. Feature selection for portfolio optimization. *Ann. Oper. Res.* **2017**, *256*, 21–40. <https://doi.org/10.1007/s10479-016-2155-y>.
- [73] Nanda, S.R.; Mahanty, B.; Tiwari, M.K. Clustering Indian stock market data for portfolio management. *Expert Syst. Appl.* **2010**, *37*, 8793–8798. <https://doi.org/10.1016/j.eswa.2010.06.026>.
- [74] Bezdek, J.C.; Pal, N.R. Cluster validation with generalized Dunn’s indices. In Proceedings of the 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, Dunedin, New Zealand, 20–23 November 1995; pp. 190–193.
- [75] Navarro, M.M.; Young, M.N.; Prasetyo, Y.T.; Taylar, J.V. Stock market optimization amidst the COVID-19 pandemic: Technical analysis, K-means algorithm, and mean-variance model (TAKMV) approach. *Heliyon* **2023**, *9*, e17577.
- [76] Wu, D.; Wang, X.; Wu, S. Construction of stock portfolios based on k-means clustering of continuous trend features. *Knowl.-Based Syst.* **2022**, *252*, 109358.
- [77] Chen, R.R.; Huang, W.K.; Yeh, S.K. Particle swarm optimization approach to portfolio construction. *Intell. Syst. Account. Financ. Manag.* **2021**, *28*, 182–194.
- [78] Data, F. Complete Intraday Bundle. Available online: <https://firstratedata.com/cb/1/complete-us-stocks-index-etf-futures> (accessed on 30 April 2023).
- [79] Platanakis, E.; Urquhart, A. Should investors include bitcoin in their portfolios? A portfolio theory approach. *Br. Account. Rev.* **2020**, *52*, 100837.
- [80] LOUIS, F.R.B.O.S. 3-Month Treasury Bill Secondary Market Rate, Discount Basis (TB3MS). Available online: <https://fred.stlouisfed.org/series/TB3MS#0> (accessed on 30 May 2023).
- [81] Top 25 Stocks in the S&P 500 By Index Weight for March 2025. Available online: <https://www.investopedia.com/best-25-sp500-stocks-8550793> (accessed on 9 April 2024).
- [82] Elton, E.J. Presidential Address: Expected Return, Realized Return, and Asset Pricing Tests. *J. Financ.* **1999**, *54*, 1199–1220. <https://doi.org/10.1111/0022-1082.00144>.
- [83] Daily Returns Meaning. Available online: <https://www.stockopedia.com/ratios/daily-volatility-12000/> (accessed on 5th October 2024).
- [84] Peng, J.; Hahn, J.; Huang, K.W. Handling missing values in information systems research: A review of methods and assumptions. *Inf. Syst. Res.* **2023**, *34*, 5–26.
- [85] Pratama, I.; Permanasari, A.E.; Ardiyanto, I.; Indrayani, R. A review of missing values handling methods on time-series data. In Proceedings of the 2016 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 24–27 October 2016; pp. 1–6.
- [86] Uddin, A.; Tao, X.; Chou, C.C.; Yu, D. Are missing values important for earnings forecasts? A machine learning perspective. *Quant. Financ.* **2022**, *22*, 1113–1132.

- [87] Kofman, P.; Sharpe, I.G. Using multiple imputation in the analysis of incomplete observations in finance. *J. Financ. Econom.* **2003**, *1*, 216–249.
- [88] Chen, A.Y.; McCoy, J. Missing values handling for machine learning portfolios. *J. Financ. Econ.* **2024**, *155*, 103815.
- [89] Wang, C.H.; Zeng, Y.; Yuan, J. Two-stage stock portfolio optimization based on AI-powered price prediction and mean-CVaR models. *Expert Syst. Appl.* **2024**, *255*, 124555.
- [90] Ojha, A.; Saxena, V. Understanding stock market trends using simple moving average (SMA) and exponential moving average (EMA) indicators. In Proceedings of the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 14–16 September 2023; Volume 6, pp. 1931–1935.
- [91] Time Series and Moving Averages. Available online: <https://www.accaglobal.com/ie/en/student/exam-support-resources/fundamentals-exams-study-resources/f5/technical-articles/time-series.html#:~:text=The%20first%20four%20observations%20are,together%20and%20dividing%20by%20two> (accessed on 20th October 2024).
- [92] Amal, M.A.; Napitupulu, H.; Sukono. Particle Swarm Optimization Algorithm for Determining Global Optima of Investment Portfolio Weight Using Mean-Value-at-Risk Model in Banking Sector Stocks. *Mathematics* **2024**, *12*, 3920.
- [93] Xu, F.; Chen, W.; Yang, L. Improved Particle Swarm Optimization for Realistic Portfolio Selection. In Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007), Qingdao, China, 30 July–1 August 2007; Volume 1, pp. 185–190. <https://doi.org/10.1109/SNPD.2007.375>.
- [94] Črepinšek, M.; Liu, S.H.; Mernik, M. Exploration and exploitation in evolutionary algorithms: A survey. *ACM Comput. Surv. (CSUR)* **2013**, *45*, 1–33.
- [95] Freitas, D.; Lopes, L.G.; Morgado-Dias, F. Particle swarm optimisation: A historical review up to the current developments. *Entropy* **2020**, *22*, 362.
- [96] Mean-Variance Optimization. Available online: <https://pyportfolioopt.readthedocs.io/en/latest/MeanVariance.html> (accessed on 20 July 2023).
- [97] Jensen, T.I.; Kelly, B.T.; Malamud, S.; Pedersen, L.H. Machine Learning and the Implementable Efficient Frontier. Swiss Finance Institute Research Paper. 2024. Available online: <https://ssrn.com/abstract=4187217> (accessed on 20th March 2025).
- [98] Merton, R.C. An analytic derivation of the efficient portfolio frontier. *J. Financ. Quant. Anal.* **1972**, *7*, 1851–1872.
- [99] General Efficient Frontier. Available online: <https://pyportfolioopt.readthedocs.io/en/latest/GeneralEfficientFrontier.html> (accessed on 15th July 2023).
- [100] Lorenzo, L.; Arroyo, J. Analysis of the cryptocurrency market using different prototype-based clustering techniques. *Financ. Innov.* **2022**, *8*, 7. <https://doi.org/10.1186/s40854-021-00310-9>.
- [101] Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

- [102] Thorndike, R.L. Who belongs in the family? *Psychometrika* **1953**, *18*, 267–276. <https://doi.org/10.1007/BF02289263>.
- [103] Shutaywi, M.; Kachouie, N.N. Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy* **2021**, *23*, 759.
- [104] sklearn extra K-Medoids. Available online: https://scikit-learn-extra.readthedocs.io/en/stable/generated/sklearn_extra.cluster.KMedoids.html (accessed on 9th October 2024).
- [105] Understanding Small-Cap and Big-Cap Stocks. Available online: <https://www.investopedia.com/insights/understanding-small-and-big-cap-stocks/> (accessed on 12th October 2024).
- [106] Sanderson, R.; Lumpkin-Sowers, N.L. Buy and hold in the new age of stock market volatility: A story about ETFs. *Int. J. Financ. Stud.* **2018**, *6*, 79.
- [107] Evans, J.L. The random walk hypothesis, portfolio analysis and the buy-and-hold criterion. *J. Financ. Quant. Anal.* **1968**, *3*, 327–342.
- [108] Paiva, F.D.; Cardoso, R.T.N.; Hanaoka, G.P.; Duarte, W.M. Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Syst. Appl.* **2019**, *115*, 635–655.
- [109] Nzokem, A.; Maposa, D. Bitcoin versus s&p 500 index: Return and risk analysis. *Math. Comput. Appl.* **2024**, *29*, 44.
- [110] Caferra, R.; Vidal-Tomás, D. Who raised from the abyss? A comparison between cryptocurrency and stock market dynamics during the COVID-19 pandemic. *Financ. Res. Lett.* **2021**, *43*, 101954.
- [111] Brini, A.; Lenz, J. A comparison of cryptocurrency volatility-benchmarking new and mature asset classes. *Financ. Innov.* **2024**, *10*, 122.
- [112] Alonso-Monsalve, S.; Suárez-Cetrulo, A.L.; Cervantes, A.; Quintana, D. Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Syst. Appl.* **2020**, *149*, 113250.
- [113] Source code for Efficient Frontier class in Python. Available online: https://pyportfolioopt.readthedocs.io/en/latest/_modules/pyportfolioopt/efficient_frontier/efficient_frontier.html (accessed on 9th October 2024).
- [114] Aljinović, Z.; Marasović, B.; Šestanović, T. Cryptocurrency portfolio selection—a multicriteria approach. *Mathematics* **2021**, *9*, 1677.
- [115] CoinMarketCap—Cryptocurrency Prices by Market Cap. Available online: <https://coinmarketcap.com>. (accessed on 7 July 2024).
- [116] The 100 Largest Companies in the World by Market Capitalization in 2024. Available online: <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-capitalization/> (accessed on 12th January 2025).
- [117] Kim, Y.B.; Kim, J.G.; Kim, W.; Im, J.H.; Kim, T.H.; Kang, S.J.; Kim, C.H. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS ONE* **2016**, *11*, e0161197.