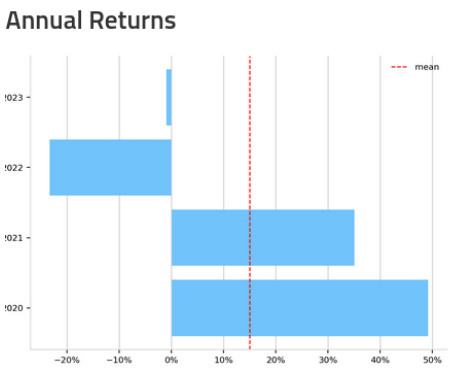
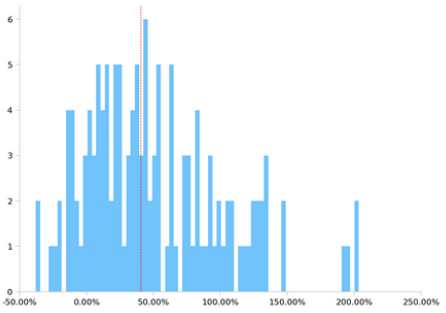
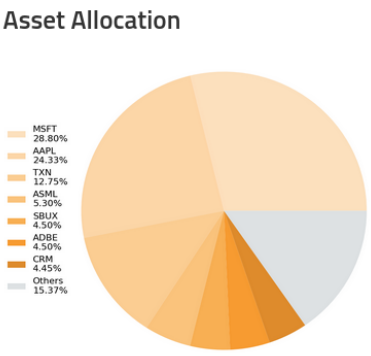
Democratizing Index Tracking for Small Investors in Europe: A Meta-Learning Approach for Sparse Portfolio Optimization

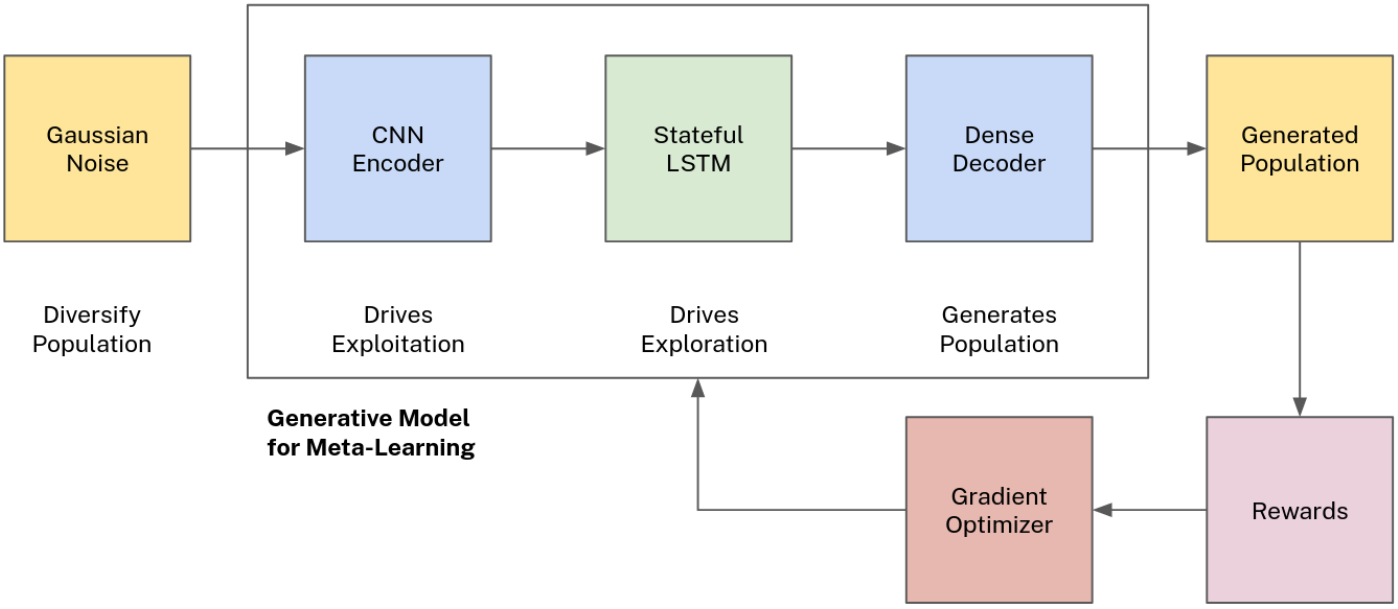
Investing in stocks is a common way for individuals to grow their wealth and diversify their investment portfolio. One popular method for investing in stocks is to buy an exchange-traded fund (ETF) or mutual fund that offers an actively managed index fund, which tracks a stock market index. However, for investors living in Europe, many of these funds are not available due to UCITS regulations. As an alternative, small investors may consider creating their own portfolio of stocks based on an index. One challenge with this approach is that most stock brokerages do not support fractional investing, meaning that an investor would need a significant amount of money to accurately duplicate the constituents of an index. One solution to this problem is known as sparse index tracking, where an investor selects a sparse portfolio that acts most similar to a given index. However, selecting the optimum sparse portfolio from thousands of stocks is a large-scale non-convex optimization problem.



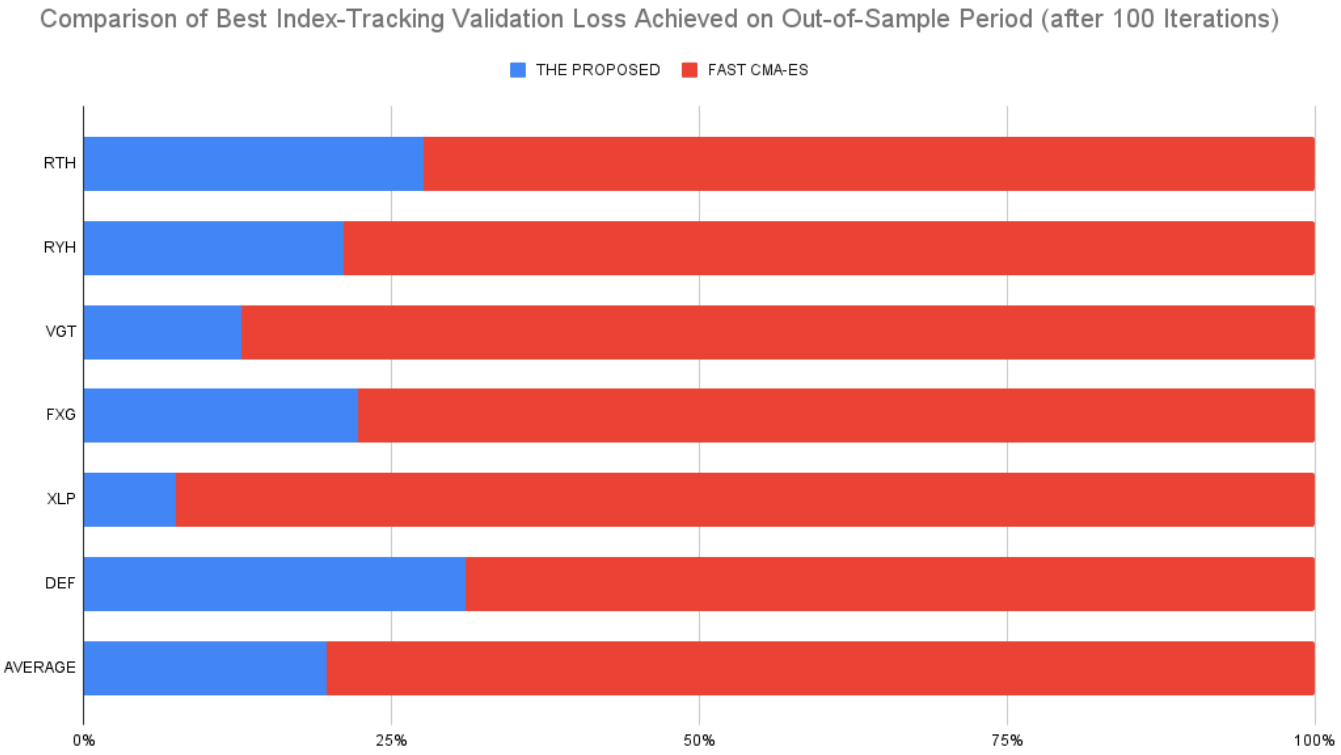




In this work, we proposed a novel population-based optimization method based on a Deep Generative Neural Network trained with Policy Gradient to learn how to sample high-quality solutions. Then, we have compared the proposed meta-learning method to the state-of-the-art heuristic algorithm for population-based black-box optimization called Fast CMA-ES, in the context of sparse index tracking. In both cases, we used the SparseMax activation function to obtain sparsified portfolio weights from the sampled populations of parameters. The optimization target for both methods was to minimize the mean squared differences between the logarithmic returns of the portfolio candidate and the index. To make it more challenging, we set the excess returns of the portfolio to zero before calculating the differences, so that portfolios are only penalized for their negative deviations from the index returns.



The experiments show that the proposed meta-learning method can converge to better solutions more efficiently when tested on an out-of-sample period. The results also indicate that most top-performing indexes can be tracked accurately with a portfolio value of at least €10,000 invested in ten or more stocks. The proposed method and the Fast CMA-ES algorithm were implemented on GPU using the PyTorch framework, and their implementations are open-sourced on GitHub. The dataset, including 14 years of daily returns from 2458 stocks from the NASDAQ and NYSE stock exchanges, is also shared to enable the reproducibility of the experiments. The QuantConnect script for the constant rebalancing of the optimized portfolio, as well as the blind back-test from that platform, are also included.



Overall, our work aims to democratize index tracking for small investors in European Union by providing a more accessible and efficient method for selecting sparse portfolios that track the performance of indexes. By using open-sourcing the implementation and dataset, we hope to enable other researchers and practitioners to build upon and further improve upon our proposed method.