Stock Market Prediction and Portfolio Composition Using a Hybrid Approach Combined with Self-Adaptive Evolutionary Algorithms

Rodrigo Manuel Oliveira de Gouvêa Lopes de Almeida
Institutu Superior Técnico
Lisbon, Portugal
rodrigolopesalmeida@tecnico.ulisboa.pt

Abstract—This work presents a new approach to maximize financial market investment returns. It incorporates two Evolutionary Algorithms (EAs) combined with fundamental and technical investment strategies. The first EA (simple) maintains its evolutionary parameters static during evolution. The second (self-adaptive) introduces the variation operators’ parameters’ values in the representation for them to evolve. The EA is responsible for optimizing the weight that financial ratios from the F-Score have on composing static/dynamic portfolios. Furthermore, it is also responsible for defining the importance that selected technical indicators have on revealing the best timing for market positions placement. A fundamental and a technical case study was created employing companies from the Standard & Poor’s 500 (SP500). These were trained/tested in a sliding window scheme between 01/2012 and 12/2018. Results showed that both case studies surpassed the SP500 returns, performing their best results using a self-adaptive EA combined with a static portfolio and a sliding window of 2 years of train/test. On the one hand, the technical case study showed better results in “bear markets” since it predicted some market declines. Its best subtest achieved returns on average 2.2x and in its best 3.5x higher than the benchmark. Its Sharpe Ratio achieved, on average, 4.9x and in its best 9x higher results than the benchmark. On the other hand, the fundamental case study displayed better performance in the “bull market”, achieving high market prices. Its best subtest achieved returns on average 2.4x and in its best 3.2x higher than the benchmark. Its Sharpe Ratio achieved, on average, 4.4x and in its best 6.5x higher results than the benchmark.

Index Terms—Evolutionary Algorithms; Self-adaptive Evolutionary Algorithms; Technical Analysis; Fundamental Analysis; Technical Indicators; F-Score; S&P500

I. INTRODUCTION

Financial markets have been around for more than four centuries. In 1602 the first “modern” securities market was founded in the Netherlands not long after the establishment of the Dutch East India Company. This creation was the result of the necessity of exchanging commodities between traders. Since the creation of the Amsterdam Stock Exchange many have followed its steps, and today a financial market is available in most nations of the modern world.

For a long time, markets have been a territory of interest not only to traders but also for researchers. The trends of the market, the price unpredictability, and the amount of money that generates created such interest that multiple theories and studies to comprehend the markets have been developed.

In the Theory of Efficient Markets (EMT), Fama et al. (1970) state that market prices reflect all the available information and that the market reacts rapidly to variations in the world [1]. This hypothesis additionally affirms that stocks trade at fair value. With that in mind, it is inconceivable for the investors to purchase undervalued stocks and make a profit out of them (beat the market). Even though this hypothesis has substantial valid premises, there is also the opposite side of supporters who do not agree with it and have proposed numerous strategies to vanquish the market.

Two main strategies attempt to beat the market: the Fundamental and the Technical. On the one hand, the fundamental strategy bases itself on financial statements and applies this economic data to build portfolios [2]. On the other hand, the technical strategy uses past data (prices/indicators) and tries to mine patterns like trends that may prompt a superior sign to where the market is moving [3].

With the expansion and study of the financial markets, numerous models were developed, and much data was generated. Many brokers also started offering more technical information. With the power of computers, many saw the possibility of processing all this information in a short time to help with trading decisions. From that point forward, multiple researchers from the most different areas have developed systems and hypotheses that enable the traders to have more power in time to make a trade. A particular area that has brought much innovation is the Artificial Intelligence (AI) domain that, in models such as Evolutionary Computation (EC) and Neural Networks (NN), has demonstrated potential on the research for forecasting models [4]–[8].

In this work, a new approach to portfolio composition is proposed, combining Evolutionary Computation with multiple indicators, ratios, and proposed theories. To solve the problem of which stocks to choose for a portfolio composition, fundamental indicators and ratios are suggested in combination with an EA. The time and type of position also have incredible significance, so the Evolutionary Algorithm helps, with the support of technical indicators, to perceive the best time to make the position in the markets.

The Evolutionary Algorithm relies on creating populations that based, on their fitness, evolves the best individuals and
removes the ones who got the worst results. With some evolution time, the population winds up showing signs of improvement and can advance the procedures of portfolio arrangement and market exchanging.

There are three main contributions in this paper. The first is the implementation of a portfolio composition system combined with F-Score for ranking companies to take part in static/dynamic portfolios. The second contribution is the development of EAs (with one being self-adaptive) that, combined with a fundamental and technical investment strategy, choose the best weights for portfolio composition variables and technical analysis indicators. Finally, the last contribution is the creation of a sliding window scheme to understand the best combination of years, for training and testing trading algorithms.

This work is structured as follows: Chapter II addresses, in the first section, the fundamental knowledge and concepts both in the markets field as in the Artificial Intelligence area. In the second part, it describes the related work and analyzes what ideas and concepts can be taken to use for further development. Chapter III presents the proposed architecture, its functionalities, and an in-depth analysis is performed for most developed modules. Chapter IV presents and examines the case studies developed in this work according to selected validation metrics. Chapter V presents the conclusions of this paper and suggestions for future work.

II. RELATED WORK

This chapter provides, in a first section, fundamental concepts on Equity Financial Markets, Market Forecasting, as well as the tools used to perform it and Artificial Intelligence with a particular focus on Evolutionary Algorithms. In the second part, the related work is presented in Market Forecasting and Evolutionary Computation.

A. Financial Markets

The concept of the financial market covers multiple types of marketplaces where several sorts of trading occur. The financial market is the name given to a place where buyers and sellers meet to trade assets. The investors or “floor traders” buy and sell assets in the exchanges to make the most profit they can. These decisions of buying and selling are based on expectations about future prices, and these, in turn, are conditional on present buy and sell decisions.

Equities are usually known as shares or stocks that represent partial ownership of a company and also work as financial instruments. Shares are formed initially when a corporation is established. The capital market is a general term that includes the stock market and other venues for trading financial products and long-term debt. The stock market allows banking institutions and investors to exchange stocks, either publicly or privately, to match capital savers with capital needers and to raise capital. Following the public listing of a company (Initial Public Offering (IPO)), the shares trade on stock exchanges and their valuation is influenced by supply and demand, which is determined by the underlying fundamentals of the macroeconomic business factors like the interest rates, and market sentiment. The profit of shareholders is a result of both the dividends paid to them from the companies profits and of any actions in the share price (capital growth).  

B. Market Forecasting

Market Forecasting is defined as the process of determining the direction in which the price is likely to move. Forecasting can be approached in two ways: using Technical Analysis or Fundamental Analysis. While technical analysis concentrates on the examination of market activity, fundamental analysis centers around the financial powers of supply and demand that cause the price to vary.

1) Fundamental Analysis: It is believed that stocks have a real “fundamental” value, different from their current market price. The fundamental value of a stock should be defined concerning the earning power of the assets or concerning the fundamental value of other stocks. Throughout time, the market price of a stock should tend towards its fundamental value, and when it happens, the analysis of fundamental values provide a useful guide to investors. Fundamental Indicators are a crucial part of the fundamental analysis. These tools provide the analyst with better methods for understanding the financial statements and formulate the appropriate conclusions.

The F-Score is a fundamental scoring system that ranges from 0-9 (9 being the best), aiming to classify how strong the financial position of a company is. To make this classification, Piotroski et al. (2012) proposed an aggregate signal that results from the addition of specific individual binary signals. These signals are financial ratios inferred from reports such as the financial reports.

\[ F\_SCORE = F\_ROA + F\_\Delta\_ROA + F\_CFO + F\_\Delta\_MARGIN + F\_\Delta\_TURN + F\_\Delta\_LEVER + F\_\Delta\_LIQUID + EQ\_OFFER \]  (1)

The binary values are always given, taking into account the result of the ratios. The ratios used in F-Score are described below:

- **Return On Assets (ROA)** allows the analyst to understand how profitable a company is.
- **Cash Flow Operations (CFO)** allows the analyst to understand the easiness that a company has to transform sales into cash.
- **Accrual** is the records of the revenues and expenses regarding a company during a specific period.
- **Gross Margin (MARGIN)** is a percentage that expresses how profitable the company’s operations are. It measures how much revenue from sales the organization holds after all the costs associated with providing a service or making a product are deducted.
- **Asset Turnover (TURN)** compares the sales value of a company to its average sales value. It is a clear way to
measure the efficiency of how the company is managing its assets to create value.

- The **change in leverage (LEV)** allows the analyst to have oversight over the changes in the company debt related issues.
- The **Liquid** variable measures the Current Ratio (CR) in comparison to the previous year’s current ratio. This result is a good way for the analyst to examine the capacity that the company has on paying debts.
- **(EQ OFFER)** tells the analyst if there is a change in the number of shares since the previous year

2) **Technical Analysis:** While fundamental analysis uses the information that’s available in the financial statements to forecast prices, technical analysis is based on finding patterns in data (as volume, open interest, sentiment measures, or others) available from databases [3]. During the technical analysis, the analyst looks at past data and searches for patterns like trends or regular cycles that could lead to correct market predictions. After the pattern extraction, it is possible to apply them and choose when to buy and sell stocks accordingly to the rules deduced from the data [10]. Technical analysis employs multiple types of tools such as: Support and Resistance, Trends, Technical Indicators and Market Indicators.

**Support and Resistance** represent moments where the buyers/sellers prevent the price rate to go below/above a certain threshold. This type of barrier allows traders to talk about price levels. **Trends** represent a fluctuation in price that makes prices follow a specific direction (trend-line). **Technical Indicators** are mathematical calculations on stock price and volume that enable the trader to anticipate price changes. While **Technical Indicators** provide a glimpse of how a stock is behaving, **Market Indicators** presents information on how a specific market is performing [14].

**C. Artificial Intelligence**

Artificial Intelligence is a broad scientific field of study. To best define it, we can divide it into its four major goals: thinking humanly, thinking rationally, acting humanly, and acting rationally. The first two are concerned with the thought process and reasoning, whereas the other two address behavior. The first and third measure success in terms of closeness to human performance, whereas the second and forth, measure toward an ideal performance measure denominated rationality [15].

AI separates into very distinct categories: Symbolic, Deep Structured Learning, Bayesian Networks, and Evolutionary Algorithms.

**Evolutionary Algorithms** underlying idea is to follow the concept of evolution. In these algorithms, a population is generated in a particular environment with a limited amount of supplies. The competition to get those resources creates a natural selection that, in turn, produces an increase in population fitness.

Given an optimization problem, multiple candidate solutions are generated and evaluated under a specific quality function. This function’s job (fitness function) is to evaluate the candidates. After the evaluation, the algorithm ranks the solutions from best to worst so that the best-ranked individuals get to breed and create the next generation. In order to create the new generation, variation operators (reproduction/mutation) are applied to the best candidates. After breeding, the offspring is evaluated with the fitness function and subsequently competes with the other solutions for selection.

This type of mechanisms allows the species to evolve in the direction of the best individual and also to have some randomness associated that enables new kinds of features to be tested in the environment. This process repeats itself until the best solution (optimization) is attained or until the stop criteria set previously is reached [16], [17].

**D. State of the Art**

In this section, multiple works in the financial market’s prediction and Evolutionary Computation area are presented and explained. At the end of the section, a table is presented, featuring the different algorithms and results obtained in the works introduced.

1) **Works on Financial Markets Prediction:** A system showing good results contradicting the SP500 and buy-and-hold on the losses by profiting in a down-trending is developed using genetic algorithms and an Echo State Network (ESN) for technical analysis parameter enhancing [6].

Another work that investigated the power of NN in financial markets showed that the average returns from NN are above the overall market average [18].

In order to predict the highest and lowest stock prices of each day, an Artificial Neural Network (ANN) using a Multilayer Feed-Forward Neural Network (MLP) was trained. The results show that the trading system can double the initial capital of the investor in the tested period, revealing to be a system with potential within the dataset presented [19].

A new approach is developed to automatically manage a portfolio by using a Genetic Algorithm (GA) conjugated with technical analysis [4]. The results prove that this method is successful, and the combination between the technical indicators and the GA can be very fruitful.

On another study, a Multi-Objective Genetic Algorithm (MOEA) to create and manage a stock portfolio by implementing a fundamental and technical approach is proposed [5]. The results show that the increase in the number of fundamental indicators make a positive impact on the precision of the simulation results and also in the MOEA.

A technical strategy optimizer combining the Volatility Index (VIX) indicator with a MOEA to predict the future tendency of assets price is developed. This work showed that the system outperforms market indexes and reduce investment risk. It is also observable that the returns increase when the risk rises and that the VIX indicator avoids several falls in the stock market by reducing the negative turns. [7].

1 Given what it knows, a system is rational if it can reason and does “the right thing.”
A hybrid approach is proposed using F-Score and G-Score among with a momentum strategy incorporated with past technical information. The results show that the hybrid system outperforms the moment strategy with higher returns.

2) Works on Evolutionary Algorithms: Genetic algorithms are a subtype of Evolutionary Algorithms used in optimization problems based on the theory of evolution and natural selection. With this in mind, GA seems a particularly exciting methodology to apply in the most different areas.

Genetic algorithms are proposed to improve technical analysis by generating a combination of parameters that identify the optimal reversal points of financial trends. By using GA as the enhancer of the technical indicators and combining it with (ESN), this work brings a new perspective on how it is possible to use genetic algorithms on stock trading systems.

In a typical GA, it is common to create a “study population” to measure its performance on doing specific tasks and to comprehend how the evolutionary steps result in creating a better population. With that as foundation, random types of individuals called “classifier equations” composed of an array of weights, that symbolize the value given to each technical rule are created to manage a financial portfolio by optimizing its technical indicators.

Usually, each candidate solution is assigned to a variety of chromosomes to stimulate the creation of unique individuals, providing more complete results. This work utilizes three different chromosomes to represent investment models with real constraints similar to the ones encountered by portfolio managers. Each individual of the population is composed of a sequence of values (genes). The chromosome divides itself into two parts. The first group is the financial ratio weights and the second is the trading parameters. Three types of chromosomes are used to have more broad results and to explore more on the GA approach. The algorithms used have two objectives: the Return and the Variance of the Returns. The system chooses the weights for each fundamental indicator and determines the values of the technical indicators. In this study, fitness is calculated using accumulated return and variance. Besides that, the algorithm selects the chromosomes for reproduction and applies the methods of crossover and mutation to create new elements.

III. Architecture

The designed system attends two significant problems: The selection of good fundamental companies for portfolio composition and the picking of the best time to take positions in the market. The approach to solving the given problems is based on Artificial Intelligence, more specifically in the Evolutionary Algorithms branch. On the one hand, the evolutionary process selects the best companies for the portfolio by providing the most suitable weights to give the appropriate importance to each ratio of the F-Score. On the other hand, it provides weights for the technical indicators in order to change their importance and to get better timing decisions according to how the market is behaving.

To accomplish the goals of the proposed system, it was necessary to design a framework that could handle all the steps required. The various steps are divided into different modules, and each module has a specific task that needs to achieve so that the system can be fully functional. In it is possible to see the full system architecture, the various modules that constitute it, and its data flow.

In order to better understand the architecture, a system flow diagram was created. The flow chart represents the interactions between modules and the way the developed system works. The colors of each module presented correspond to the colors of the architecture module to whom they belong.

A. Data Preparation

The Data Preparation Module handles three main duties: Data Downloading, Data Processing and Data Manipulation. Data Downloading retrieves and stores multiple required datasets from various sources. Data processing inspects, cleans and verifies the data downloaded. Data Manipulation calculates financial ratios and technical indicators from the pre-processed data for posterior use.

The financial indicators were calculated for each downloaded company using the F-Score ratios. The technical indicators were calculated using specific rules appointed to each Technical Indicator (TI) to form a score. The score varies from Very Low Score (-1), Low Score (-0.5), Neutral (0), High Score (0.5), and Very High Score (1). The indicators were selected, taking into consideration the goals to achieve and their performance in the works presented. The five indicators picked were: Relative Strength Index (RSI), Exponential Moving Average (EMA), Moving Average Crossover (MAC), Rate of Change (ROC), the Moving Average Convergence Divergence (MACD) and the VIX.

B. Evolutionary Algorithm

In this work, Evolutionary Algorithms were selected as the model of AI to implement. This type of algorithms were picked since the reviewed papers revealed that these are able to achieve promising results when combined with financial forecasting. Furthermore, the same analysis showed that EA are capable of handling the goals proposed for this work. The Evolutionary Algorithm Module is the center of this work. This module handles the creation and evolution of the system weights in order to achieve the best solution possible. In this subsection, some of the EA constraints implemented are described as the parameters used in this work.

1) Representation: In this work, the representation was developed in a modular way. There are three types of representation modules: Fundamental, Technical and Self-Adaptive.

The Fundamental Representation module is composed of nine values depicting the weights of each fundamental F-Score ratio. In figure a representation of the fundamental chromosome is presented. It is important to notice that the colors of the representation modules presented below, match the colors of the architecture modules where their values are utilized.
TABLE I: Related Work Table.

<table>
<thead>
<tr>
<th>Work</th>
<th>Date</th>
<th>Period</th>
<th>Input Data</th>
<th>Algorithm</th>
<th>Evaluation</th>
<th>Results</th>
<th>Best Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>[18]</td>
<td>2004</td>
<td>1985 to 1995</td>
<td>CompStat and Citibase</td>
<td>NN</td>
<td>AR</td>
<td>-</td>
<td>0.253</td>
</tr>
<tr>
<td>[19]</td>
<td>2009</td>
<td>May to December 2008</td>
<td>BOVESPA</td>
<td>ANN</td>
<td>AR, Maximum Drawdown and Average Number of Daily Operations</td>
<td>-</td>
<td>1892.16%</td>
</tr>
<tr>
<td>[6]</td>
<td>2011</td>
<td>Bull – 2003 to 2005</td>
<td>S&amp;P500</td>
<td>GA and ESN</td>
<td>How close the suggested trading points are to the turning points of price trends</td>
<td>41.6%</td>
<td>-</td>
</tr>
<tr>
<td>[7]</td>
<td>2015</td>
<td>2006 to 2014</td>
<td>NASDAQ, S&amp;P500, FTSE100, DAX30 and NIKEE225</td>
<td>MOEA with Pareto Fronts</td>
<td>ROE and Variance of Returns in a Portfolio</td>
<td>AR &gt; 10%</td>
<td>-</td>
</tr>
<tr>
<td>[20]</td>
<td>2016</td>
<td>1975 to 2013</td>
<td>NYSE, AMEX, NASDAQ</td>
<td>BOS Momentum with F-Score and G-Score</td>
<td>Past six months winner and loser stocks</td>
<td>-</td>
<td>1.04% (Average Monthly Return Over the Price Momentum Strategy)</td>
</tr>
</tbody>
</table>

The Technical Representation module is composed of six weights assigned for each technical indicator score created in the Data Module. In figure 4, a representation of the technical chromosome is presented.

The Self-Adaptive Representation module is composed of two genes that dictate the probabilities of mutation $P_m$ and crossover $P_c$ variation operators parameters. This representation module was created considering one of the biggest problems in EA: Parameter tuning. Building an executable EA instance requires stipulating values for its parameters. These values determine whether it will find an optimal solution and whether it will do so efficiently. Parameter tuning is a commonly practiced approach to algorithm design, where, in most cases, values for the parameters are defined prior to the beginning of the algorithm, staying fixed throughout the whole process.

The traditional method to choose the parameters is based on conventions such as “mutation probability should be low”, “population size should be 200” and limited experimentation with distinct values. For example, considering three parameters and four values for each of them.

Nevertheless, this type of parameter tuning is not feasible. Trying all different combinations systematically is extremely time-consuming. Besides, for numerical parameters, the optimal values could lie between the points we are testing or not even be among the ones selected. This idea becomes even more discouraging if we are looking for a generally good setup that is able to operate well on a range of problems/scenarios.

In this work, the self-adaptive representation allows the EA to incorporate the variation parameter values in the candidate solutions. With this methodology, the EA evolves and adapts the variation operator parameters, hoping to achieve optimal values for efficient optimization. In figure 5, a representation of the self-adaptive chromosome is presented.

In this work, the representation modules were combined in order to achieve various types of representations. The outcome of the combinations led to the creation of four types of representations. The first representation is used for a fundamental (Buy and
The second type of representation is used for a technical investment strategy and combines the fundamental with the technical representation module. This type of representation uses its fundamental weights and ratios to create a portfolio but also the technical indicators scores and respective weights to make positions in the market.

The third and fourth types of representations used in this work are variations of the first and second presented with the addition of the self-adaptive representation module to add self-adaptive capabilities. These last types of representations are used for fundamental and technical investments, respectively.

2) Fitness Function: For the evolution process to occur an evaluation metric to distinguish the best solutions from the worst must exist.

In this work, the fitness function utilized was the Rate of Return (ROR). This metric is used to evaluate the gain/loss of some investment over time, expressed as a percentage of the investment initial cost. The higher the value of this metric, the best the individual performance is as it can obtain high returns on its investments. The formula follows:

\[
ROR = \frac{GainFromInvestment − CostOfInvestment}{CostOfInvestment} \times 100
\]

3) Variation Operators: After the parents have been selected, variation operators are employed to generate the new offspring. For the crossover procedure, the N-point crossover (with 2 points) was selected using a 100% crossover probability \((P_c)\). The mutation selected works as follows: For each gene of the chromosome, a probability/value is generated randomly. If the value generated is lower than the mutation probability \((P_m)\), then a new value is generated for this specific gene. For this operator, a probability of 20% was selected. In a self-adaptive scenario, the \((P_c)\) and \((P_m)\) are determined by the value in the chromosome.

4) Replacement: After the variation operators have completed their job its time for the old population to be replaced with the offspring. In this work, the previous generation population is almost entirely replaced with the offspring. A small percentage of the new population is reserved for the elite of the old population. Elitism was implemented in order to maintain the best solutions from previous iterations. The percentage of the population that represents the elite is 30%.

C. Portfolio Creator

The Portfolio Creator Module is where the portfolio is constructed according to the year to trade, the weights generated in the Evolutionary Algorithm and the size of the portfolio desired.

In this work, the portfolio size established was 20 companies since its essential to diversify the capital into various investments to minimize the risk exposure. The portfolio is always created, taking into account the year before entering the market. This strategy allows choosing the best companies of the year before, hoping that they will behave as well in the next year.

In order to create the portfolio, the F-Score \([I]\), along with the fundamental datasets created in the Data Preparation Module, are used to score the financial ratios. The formula awards the ratios with binary values according to its rules. After that, each score is multiplied by its corresponding weight from the chromosome to give the necessary importance to each ratio. After the F-Score \([I]\) has been calculated for all the companies, they get ranked from highest to lowest score. Subsequently, the companies with the best scores are selected to be part of the portfolio.

D. Trading Simulator

The Trading Simulator is a close to the real-life trading environment responsible for training/testing the chromosomes and portfolios created. This simulator runs through every data point in sequence and for each allows the portfolio companies to take a position. Each company can only have one active position meaning that if the company is already bought, the only positions the system is allowed to execute are to maintain the investment or to sell it. Furthermore, the simulator invests the same amount of money into each company. At the end of each year, all the stocks unsold are sold, and depending on the settings selected, a new portfolio might be created for the upcoming year. When the simulation finishes, the results are sent to the Statistics Module to calculate its return that serves as fitness for the chromosome and portfolio used.

The Trading Simulator is responsible for two main points:

1) Dynamic F-Score: The Dynamic F-Score is the name given to a feature developed to update the portfolio each year. The goal of this innovation relies on getting the best-performing companies from the year \(y - 1\) ready to trade in the year \(y\), hoping that their fundamentals will enable good trading performance. When this feature is enabled, the Trading Simulator Module “asks” the Portfolio Creator Module to build a new portfolio for the upcoming year.

2) Investment Strategies: There are multiple approaches to the problem of when to take positions in the market. In this work, two types of strategies were implemented: Buy and Hold and Technical Investment.

i. The Buy and Hold strategy utilizes the portfolio created and buys stocks for each company on the first day of trading and sells them all on the last day of trading.

ii. The Technical Investment utilizes the portfolio created and takes positions according to a weighted average

![Pm Pc](image-url)
between the technical indicators scores and its chromosomes weights.

The result is then interpreted in the following way:
- If the result is $> 0.5$, then it represents a Buy position.
- If the result is $< -0.5$, then it represents a Sell position.
- Any other result represents a Maintain position.

E. Statistics

The Statistics Module has two main goals: Calculate the fitness values for the candidate solutions using the data frame received from the Trading Simulator Module and the fitness function (2) but also calculate statistics for the simulations in order to validate and evaluate the system performance. In this work, metrics such as Maximum Drawdown (MDD), Risk-Return Ratio (RRR), Sharpe Ratio, Sortino Ratio, and the fitness function ROR were employed to evaluate the system performance. Besides other metrics such as the number of transactions, Profitable Transactions (PT), Unprofitable Transactions (UT), Maximum Profit (MaxP), Minimum Profit (MinP), Average Profit (AvgP) and Average Return (AvgR) were employed to evaluate the evolution of trading activity.

IV. RESULTS & DISCUSSION

This thesis comprehends two central case studies: The Technical Investment and the Fundamental Investment. Each case study subdivides itself into four specific cases:

1) With dynamic F-score
2) Without dynamic F-score
3) With dynamic F-score and self-improvement
4) Without dynamic F-score but with self-improvement

On the one hand, the first two subtests examine the influence of adopting a fundamental/technical investment strategy combined with using a portfolio that changes each year versus a portfolio that stays unchangeable throughout the whole trading period. On the other hand, the last two subtests also test the importance of using a fundamental/technical investment strategy combined with a dynamic/static portfolio. However, also study the importance of adding variation operators probabilities into the chromosome structure. The purpose of these experiments is to understand the effect that these capacities produce on trading returns.

In order to test the suggested cases, the architecture and parameters presented in chapter III were used. The subtests were submitted to a period of training and testing (out of sample) using a sliding window approach to expose the system to different environments. The developed scheme comprises blocks of one year that together comprehend data from 2012-01-01 to 2018-12-31. The multiple arrangements of these blocks compose six types of sliding windows (32,31,22,21,12 and 11). Each sliding window name represents the number of training blocks first and the number of testing after. A simple example is the window 32 that represents three years of training and two years of testing. In order to compare the results from different sliding windows, each composition includes the same four years of testing.

To obtain robust results, the trains/tests were repeated ten times each, and the metrics introduced in the Statistics chapter III were utilized as system validation/evaluation metrics. To understand how the system performance is related to a "real-life" trading method, the SP500 index prices were used as a “benchmark”.

A. Technical Investment

In this case study, the EA evolved multiple candidate solutions in order to achieve the best possible return. The candidates are composed of weights for the portfolio creation (fundamental representation module) and for the technical investment strategy (technical representation module). In the last two subcases, the candidate solutions also included the variation probabilities in their representations (self-adaptive representation module) to create a self-adaptive EA.

The first case study is characterized by achieving better results on average in the sliding window, composed of two years of training and two years of testing. Furthermore, this case study presents better results on average using a non-dynamic portfolio management strategy. Through analysis of figures 6 and the results table II presented, it is possible to conclude that the technical system was able to avoid multiple losses and present fairly consistent profits. Thanks to the selected technical indicators, the system was able to abandon the market (creating the horizontal lines seen in the multiple figures) in multiple situations that later proved to be declining moments. However, during “bull markets”, this system showed some lack of “confidence” to pursue higher returns since it would take some time to start “climbing” the market after declining moments.

Fig. 6: Technical Investment Best Subtest Results (Without Dynamic F-Score but With Self-Improvement Average Simulations).

The technical case study achieved its best average simulation without using a dynamic portfolio strategy, however, it used a self-improved EA. This simulation achieved the best ROR of all average simulations with a 35.67% ROR, 11.90% MDD, Sharpe Ratio of 1.33, RRR of 3.62, Sortino Ratio of 0.63 and a sliding window of 22. From the same subtest and the same sliding window comes the best simulation with a
B. Fundamental Investment

In this case study, the EA also evolved multiple candidate solutions in order to achieve the best return possible. The candidates are formed of weights for the portfolio creation (fundamental representation module) that is later used in the Trading Simulator Module with a Buy-and-Hold investment strategy. In the last two subcases, the candidate solutions also included the variation probabilities in their representations (self-adaptive representation module) to create a self-adaptive EA.

The second case study is characterized by achieving its best average results using two training years and one/two testing years. Additionally, this case study presents better results on average using a static portfolio management strategy. Through analysis of figures 9, 10 and the results table III presented, it is possible to conclude that the fundamental investment case study was able to achieve high profits. Thanks to the weighted F-Score scoring system and the multiple enhancements developed in this work, a simple buy and hold investment strategy proved to be highly profitable. On the one hand, not having technical indicators may have downsides since it allows the portfolio to drop multiple times during declining markets. On the other hand, during a “bull market” the portfolio is able to achieve stellar returns since it does not have a “bottleneck” to stop its profits.

The fundamental case study achieved its best average simulation in the same subtest and sliding window as the technical case study. This simulation produced the best ROR of all average simulations with 39.49%, a 17.25% MDD, Sharpe Ratio of 1.19, RRR of 2.36, and Sortino Ratio of 0.68. From the same subtest and sliding window comes the best simulation with a 52.64% ROR, MDD of 13.75%, Sharpe Ratio of 1.78, RRR of 3.82, and Sortino Ratio of 1.

In terms of self-adaptive EA, the self-dynamic subtest presents, on average, nearly equivalent results as the original test. However, the same does not happen in the static portfolio subtest, where its self-adaptive derivation shows better results in every single performance return average metric.

Through the analysis of figure 11, it is noticeable that during training, the algorithm was able to evolve its population. On the one hand, the simple dynamic EA achieved an average 55% ROR, while its self-adaptive derivation produced a nearby average value of 54% ROR as it happened in the test sessions. On the other hand, the non-dynamic subtest achieves, on average, a ROR of 53% approximately while its self-adaptive derivation produces a lower value around 50% ROR.

C. Results Comparison

In this work, both case studies achieved better returns using a self-improved EA and a static portfolio management strategy. On the one hand, the self-improved EA results do not come as a surprise. Including the variation operators’ parameters in the chromosome allow the algorithm to perform a more specialized evolution since it adapts these values according to the fitness achieved. Consequently, this feature brings a more efficient evolutionary process and, therefore, higher returns during the testing period.

On the other hand, the static portfolio results come as a negative surprise. During the elaboration of the portfolio management strategies, the dynamic approach emerged as a “logical” solution. Using the best companies from the year before trading in the following year should yield good returns. This idea demonstrated not to be always true since the static portfolio strategy overcame the dynamic strategy. The main reason for this to happen may relate to market seasonality and the world in general. While during a “bull market”, all companies tend to thrive, during a “bear market”, even the best companies decline. Not only that, but companies who thrive in specific momentum’s may not prosper in others. The reason behind it may not only be specified in financial statements but may also be connected to the state of the world. Events such as economic tensions between countries, legal agreements, wars, and scandals impact companies’ prices by creating fear or optimism in the investors. This fear/optimism then produces a significant rise or decline in the companies prices, creating unexpected market momentum’s.
Lastly, a relevant point to notice is that although the training sometimes would yield high ROR that, would not necessarily mean that their tests would be that great and vice-versa. Once again, the trading market is a very complex system that is connected to the world in general. Training a system that works flawlessly during several years may result in a good testing performance in some environments. However, it does not imply that it will be able to dominate and entirely surpass the market during testing years.

V. CONCLUSIONS

The analysis of the results obtained in this work allows concluding that Evolutionary Algorithms, combined with the stock market, represents a powerful tool when it comes to portfolio composition and technical trading. Not only that, but also the capability of using multiple data sources, extracting the critical information, transforming it into usable data to compose portfolios for trading in few seconds demonstrates significant potential in the developed system. The presented results are promising and reveal that the technical and the fundamental case studies combined with Evolutionary Algorithms are able to surpass the SP500 returns. Both case studies achieved their best results using a static portfolio management strategy and a self-adaptive EA with returns more than two times higher than the benchmark in both average cases. Besides, both subtests also achieved these results in the same sliding window composed of two years of training and two years of testing.

On the one hand, the technical case study revealed its value during “bear markets” since it is able to diminish accentuated declines. Nonetheless, during “bull markets”, the technical case study lacks “confidence” to pursue higher returns. On the other hand, the fundamental case study shows many more
TABLE III: Fundamental Investment Best Subtest Results (Without Dynamic F-Score but With Self-Improvement).

<table>
<thead>
<tr>
<th>Sliding Window</th>
<th>Epoch</th>
<th>MDD</th>
<th>RRR</th>
<th>Sharp Ratio</th>
<th>ROR</th>
<th>Sortino Ratio</th>
<th>PT</th>
<th>UT</th>
<th>MaxP</th>
<th>MinP</th>
<th>AvgP</th>
<th>AvgR</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>Average</td>
<td>16.77</td>
<td>2.05</td>
<td>0.98</td>
<td>14.37</td>
<td>0.56</td>
<td>68.25</td>
<td>31.75</td>
<td>113.67</td>
<td>1.03</td>
<td>28.59</td>
<td>13.09</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>16.79</td>
<td>2.19</td>
<td>1.20</td>
<td>40.16</td>
<td>0.69</td>
<td>75.00</td>
<td>25.00</td>
<td>110.50</td>
<td>0.61</td>
<td>29.43</td>
<td>17.29</td>
</tr>
<tr>
<td>31</td>
<td>Average</td>
<td>22.60</td>
<td>1.10</td>
<td>0.56</td>
<td>24.98</td>
<td>0.34</td>
<td>44.00</td>
<td>45.00</td>
<td>88.97</td>
<td>0.45</td>
<td>26.70</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>17.25</td>
<td>2.36</td>
<td>1.19</td>
<td>39.49</td>
<td>0.68</td>
<td>69.50</td>
<td>50.50</td>
<td>124.06</td>
<td>0.55</td>
<td>32.08</td>
<td>16.73</td>
</tr>
<tr>
<td>22</td>
<td>Average</td>
<td>21.75</td>
<td>3.62</td>
<td>1.78</td>
<td>52.64</td>
<td>0.45</td>
<td>77.50</td>
<td>22.50</td>
<td>137.61</td>
<td>0.45</td>
<td>31.69</td>
<td>22.29</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>26.09</td>
<td>0.05</td>
<td>-0.26</td>
<td>1.42</td>
<td>-0.16</td>
<td>48.00</td>
<td>52.00</td>
<td>66.95</td>
<td>0.38</td>
<td>24.51</td>
<td>1.28</td>
</tr>
<tr>
<td>21</td>
<td>Average</td>
<td>29.53</td>
<td>0.15</td>
<td>0.05</td>
<td>10.69</td>
<td>0.03</td>
<td>55.00</td>
<td>45.00</td>
<td>66.92</td>
<td>0.12</td>
<td>24.23</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>15.01</td>
<td>1.99</td>
<td>0.81</td>
<td>30.11</td>
<td>0.46</td>
<td>69.75</td>
<td>30.25</td>
<td>69.67</td>
<td>0.63</td>
<td>23.36</td>
<td>10.65</td>
</tr>
<tr>
<td>12</td>
<td>Average</td>
<td>16.42</td>
<td>2.14</td>
<td>1.17</td>
<td>38.53</td>
<td>0.66</td>
<td>70.00</td>
<td>30.00</td>
<td>110.50</td>
<td>0.61</td>
<td>28.19</td>
<td>14.67</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>24.00</td>
<td>0.27</td>
<td>-0.18</td>
<td>4.99</td>
<td>0.03</td>
<td>53.87</td>
<td>44.12</td>
<td>135.24</td>
<td>0.45</td>
<td>23.43</td>
<td>2.36</td>
</tr>
<tr>
<td>11</td>
<td>Average</td>
<td>22.18</td>
<td>1.24</td>
<td>0.59</td>
<td>27.64</td>
<td>0.38</td>
<td>56.25</td>
<td>43.75</td>
<td>427.92</td>
<td>0.45</td>
<td>31.35</td>
<td>10.67</td>
</tr>
<tr>
<td></td>
<td>Best</td>
<td>-</td>
<td>17.41</td>
<td>0.91</td>
<td>0.27</td>
<td>16.00</td>
<td>0.15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 11: Average Fundamental Training of the Evolutionary Algorithm.

declines during the “bear market”. Nevertheless, it shows the ability to “climb” the market at a fast pace achieving significant returns during “bull market” momentum’s. Overall, the proposed goals in this work were achieved, showing that it is possible to surpass SP500 using not only a fundamental approach but also a technical approach if combined with EA.

In this work, the most significant system limitation is related to the time each simulation took to complete. Although much was done to decrease time complexity, there is always room for improvement.

Some future work improvements could be replacing the developed EA with a MOEA system using two financial ratios to maximize return and reduce risk. Another improvement would be transferring the developed work into a faster version of Python to decrease the time it takes to run each simulation. Finally, in this work, the variation operators’ parameters were included in the representation for optimization. In future work, other parameters such as the “percentage of the population that takes part in elitism” could also be included in the representation for optimization.

REFERENCES


