Data Mining Approaches for Stock Risk Assessment

Amir H. Gandomi[a,*], Amirhessam Tahmassebi[b], Anke Meyer-Baese[b]

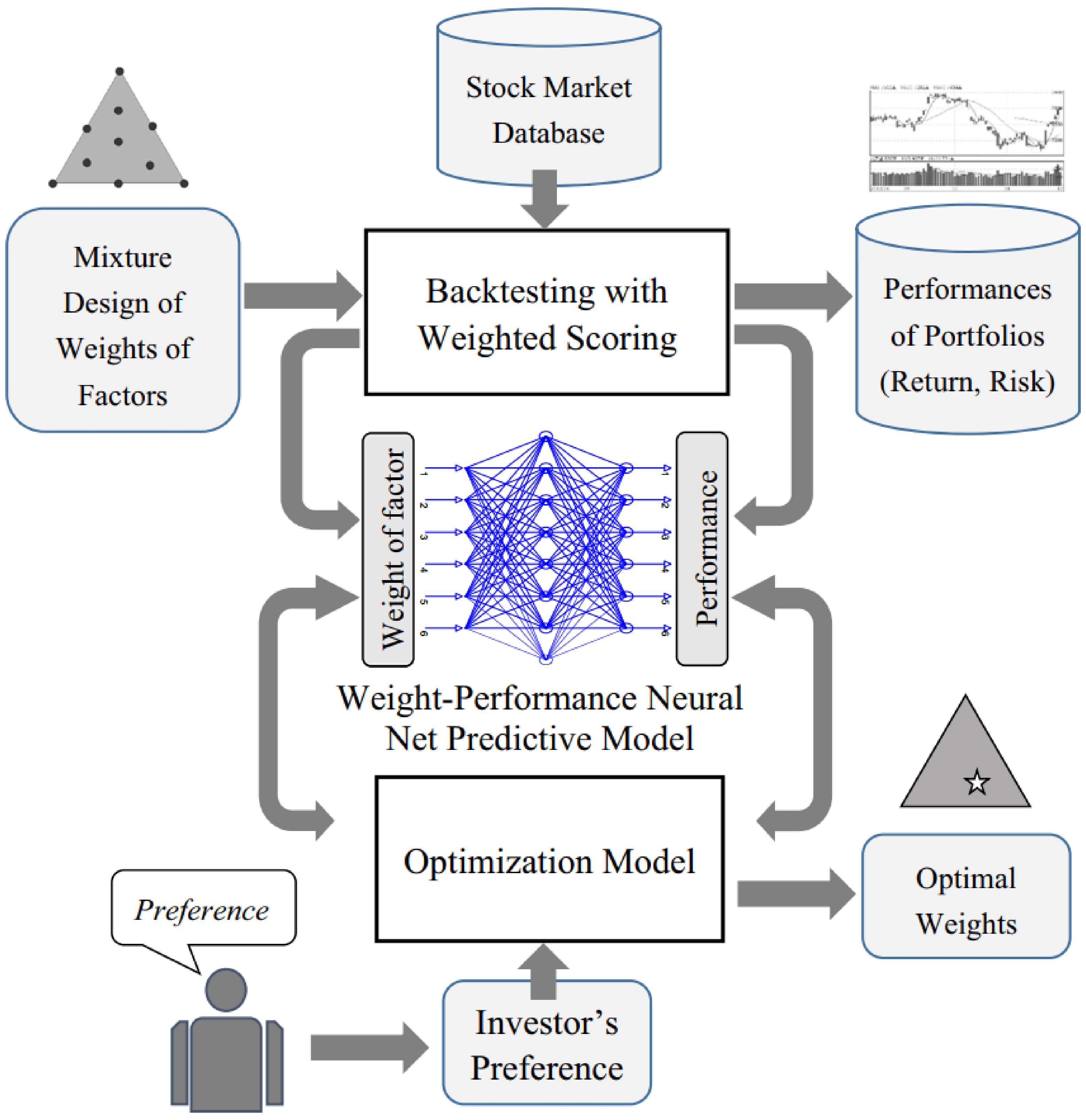
[a]School of Business, Stevens Institute of Technology, Hoboken, New Jersey 07030, USA

[b]Department of Scientific Computing, Florida State University, Tallahassee, Florida 32306-4120, USA

Abstract

The recent exponential growth of investors in stock markets brings the idea to develop predictive models to forecast the total risk of investment in stock markets. In this study, several data mining approaches including classical and evolutionary computation approaches were proposed to predict the total risk in stock investment. To develop evolutionary computation models, a multi-objective genetic programming (GP) strategy based on non-dominated sorting genetic algorithm II was employed. Here the optimization of mean-square error as the fitness measure and the subtree complexity as the complexity measure are considered as problem objectives. The GP model ran for 500 generations with 1000 population considering training/testing sets to overcome any possible over-fitting. The proposed models are developed using an S&P 500 database in a 20-year time period. These various data mining algorithms compared and evaluated. The reasonable results suggest that some of the proposed models can reach a high degree of accuracy and they can be applied to various stock databases to assess the total risk of investment. The accurate models along with stock selection decision support systems can overcome the disadvantages of weighted scoring stock selection.

Stock Selection Decision Support System



Stock Market Database

Mixture Design of Weights of Factors

Backtesting with Weighted Scoring

Performances of Portfolios (Return, Risk)

Weight of factor

Performance

Weight-Performance Neural Net Predictive Model

Optimization Model

Preference

Optimal Weights

Investor's Preference

Figure: Diagram of stock selection decision support system [1].

S&P 500 Database

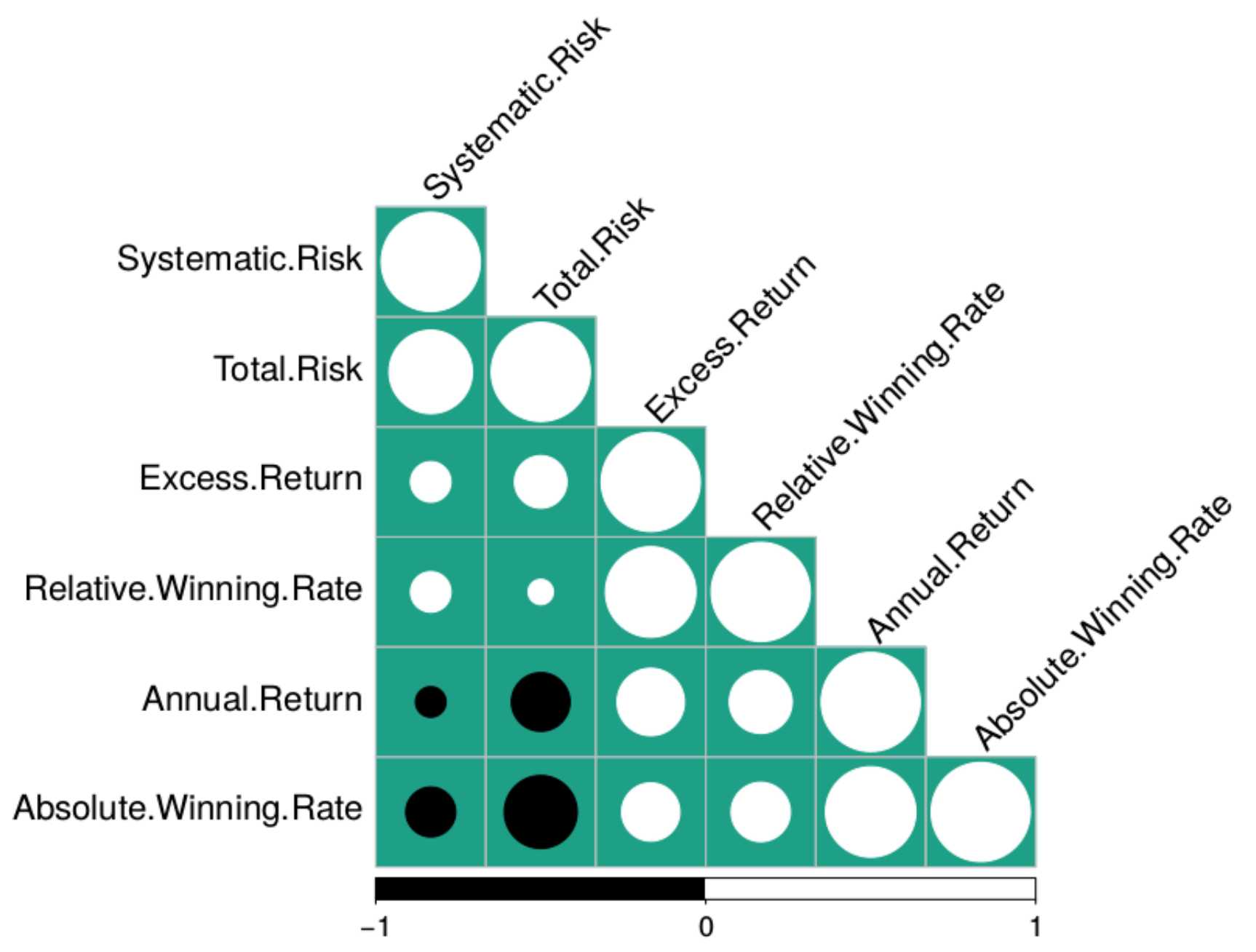


Figure: Correlation matrix illustration of the input variables along with the output variable using hierarchical clustering.

MOGP Evolution

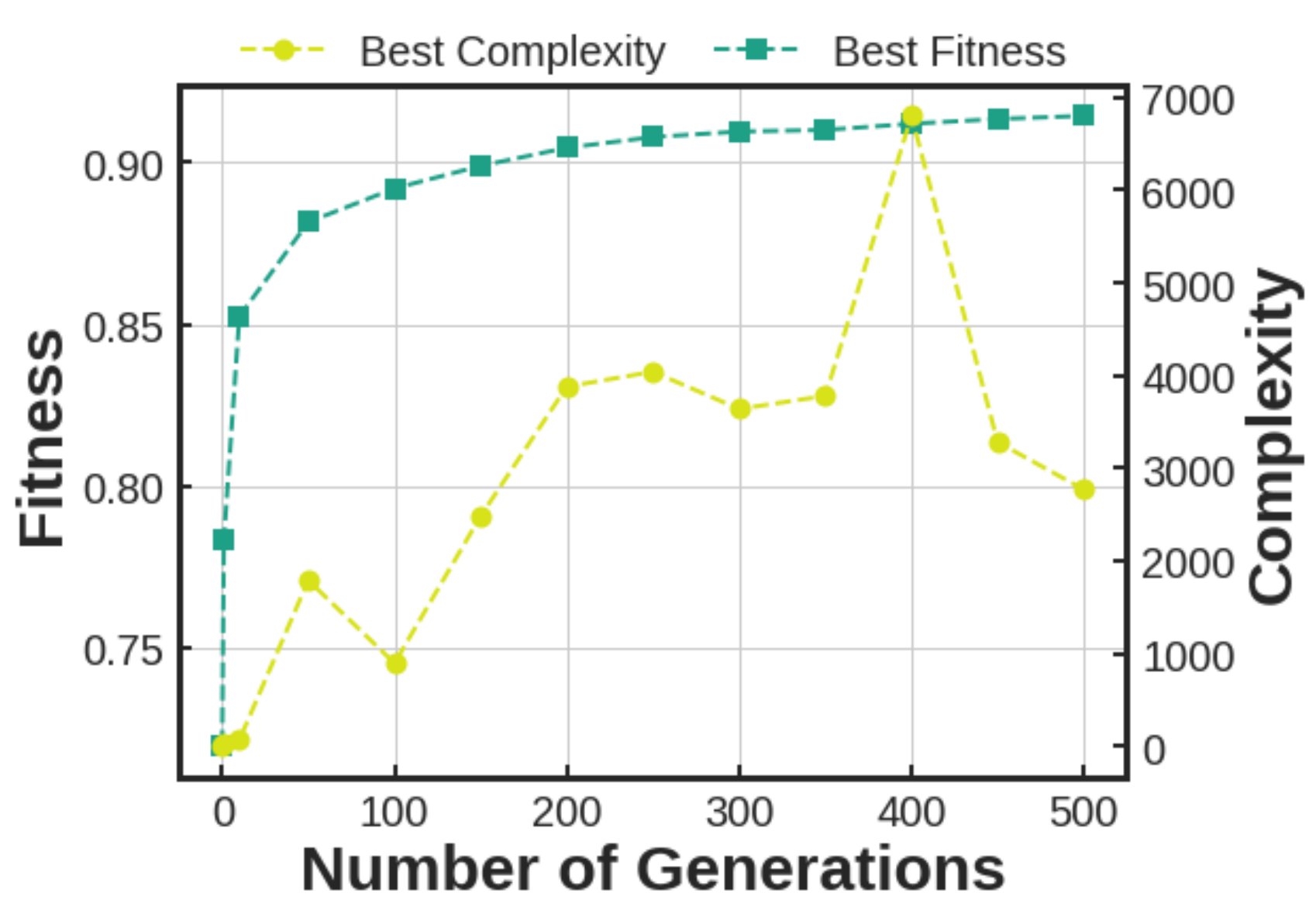
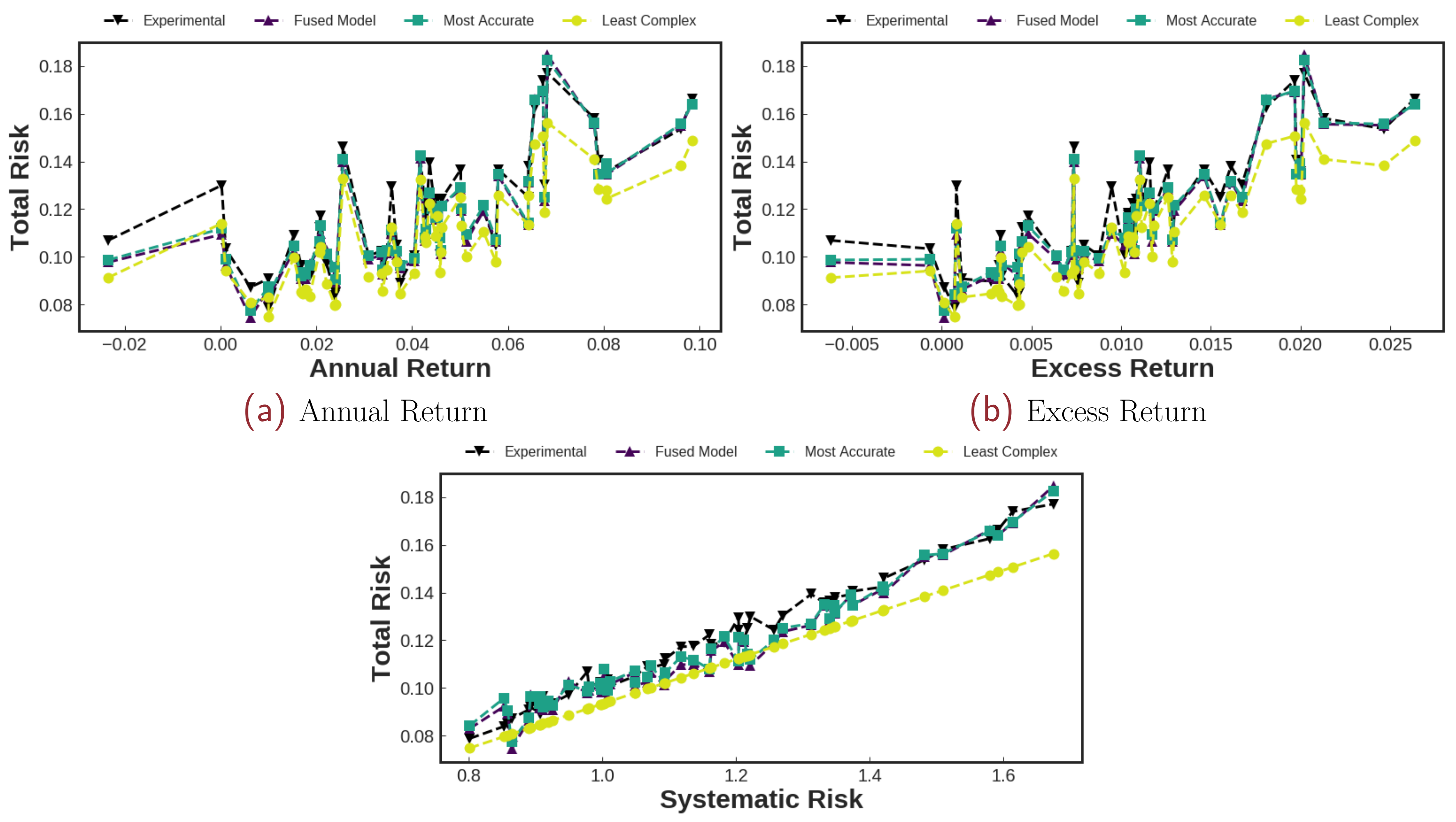


Figure: The evolution of the employed objective functions, fitness and complexity measures for the developed GP model through different numbers of generations.

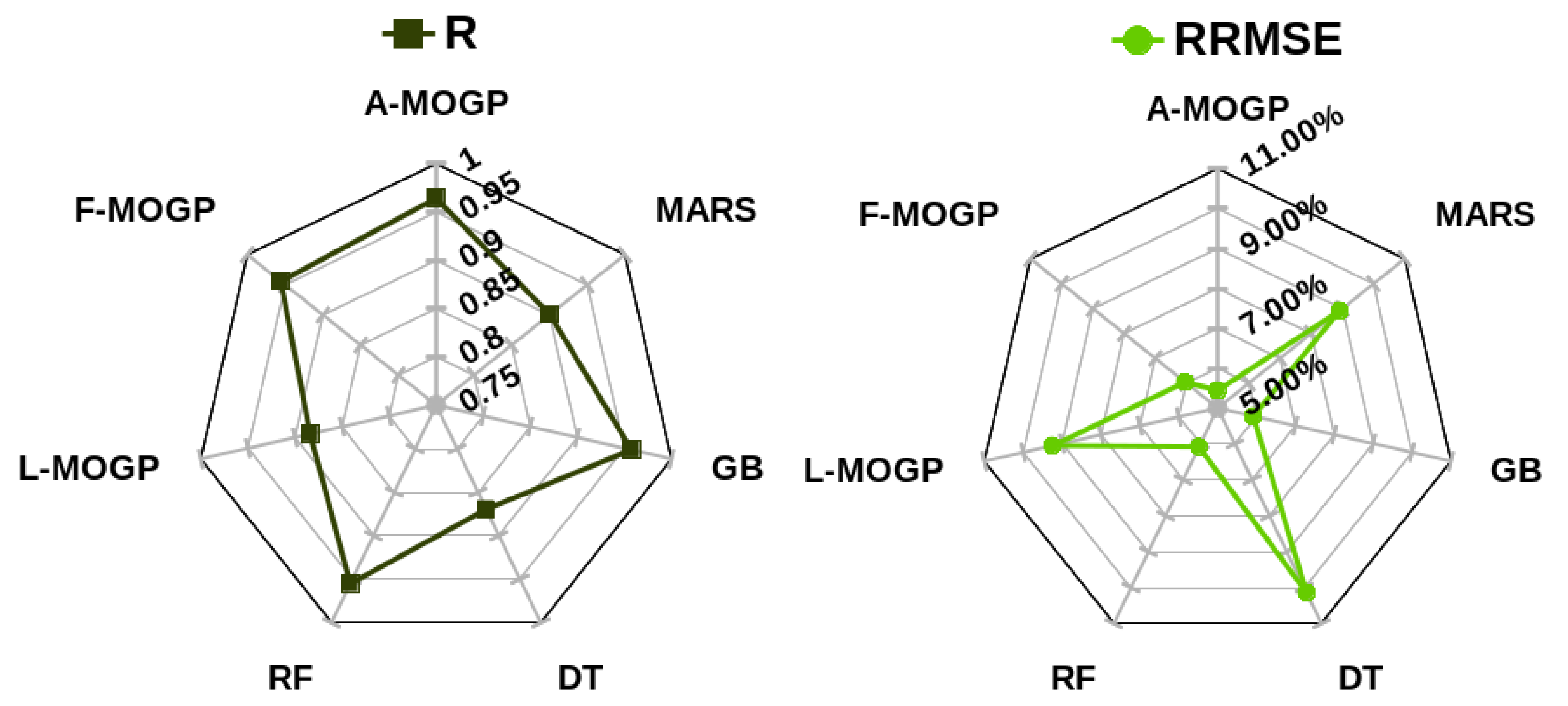
Results



(a) Annual Return (b) Excess Return

(c) Systematic Risk

Figure: An exhaustive comparison of the predicted total risk for (a) the annual return, (b) the excess return, and (c) the systematic risk using the most accurate model, the least complex model, and the fused model versus the experimental data.



(a) R (b) RRMSE

Figure: Score metrics comparison of the predicted total risk using the developed MOGP models with various machine learning algorithms.

Conclusion

This paper aims at developing classical machine learning and evolutionary computation techniques for stock prediction. In this regard, a multi-objective genetic programming (GP) strategy based on non-dominated sorting genetic algorithm II was employed. Here the optimization of mean-square error as the fitness measure and the subtree complexity as the complexity measure are considered as problem objectives. The GP model ran for 500 generations with 1000 population considering training/testing sets to overcome any possible over-fitting. Higher R values and lower $RRMSE$ values result in a more precise model.

References

[1] Yi-Cheng Liu and I-Cheng Yeh.
Using mixture design and neural networks to build stock selection decision support systems.
Neural Computing and Applications, 28(3):521–535, 2017.

Contact Information

- * Corresponding Author: Amir H. Gandomi
- Email: a.h.gandomi@stevens.edu
- URL: http://gandomi.beacon-center.org/

©Informs Business Analytics Conference, April 15-17, 2018, Baltimore, MD, USA