Value of Machine Learning in Mutual Fund Trades

Hsiu-Lang Chen, Linh H. Nguyen, and Linh X. D. Nguyen*

September 2024

Abstract

We examine the ability of an artificial neural network (ANN) to learn from the investment decisions of mutual fund managers. Our findings suggest that an ANN can use mutual fund holdings disclosure, firm characteristics, and macroeconomic variables, to learn fund managers' strategies for stock trades and provide desirable performance. Fund managers still exercise important skills that artificial intelligence (AI) cannot learn. Funds with higher levels of non-AI skills tend to generate significant and higher alphas. We conclude that investors can use ANN to identify and invest in these funds.

JEL Classification: C52, C55, G10, G11, G17, G23

Keywords: Machine Learning, Artificial Neural Networks, Mutual Funds

* This article represents the views of the authors only. Hsiu-Lang Chen, University of Illinois Chicago, Chicago, IL 60607, USA (hsiulang@uic.edu). Linh H. Nguyen, University of Nottingham, Nottingham, NG8 1BB, UK (linh.nguyen2@nottingham.ac.uk). Linh X. D. Nguyen, University of Leicester, Leicester, NG2 1RQ, UK (lxdn1@leicester.ac.uk).

1. Introduction

Interest in Artificial Intelligence (AI), especially machine learning, and its applications in finance, has grown tremendously in both academia and industry. Machine learning offers a way to accommodate high-dimensional predictor sets and flexible functional forms to uncover complex patterns and underlying relationships.

One main field of machine learning, Artificial Neural Network (ANN), has attracted considerable interest because of its outstanding performance in various financial applications, including bankruptcy prediction (Fletcher and Goss, 1993; Brockett et al., 1994), credit evaluation (Baesens et al., 2003), board director selections (Erel et al., 2021), and portfolio allocation (Lin et al., 2006; Ko and Lin, 2008). A variety of studies focus on the use of ANN in stock return prediction and stock selection. Heaton, Polson, and Witte (2016) develop a neural network for replicating the S&P 500 Index through a subset of stocks while Guresen, Kayakutlu, and Daim (2011) compare several models of ANN in forecasting the daily Nasdaq Composite Index. Ticknor (2013) proposes a three-layered feed-forward ANN with Bayesian regularization to predict daily stock prices. In a comparative analysis of machine learning methods for measuring asset risk premiums, Gu, Kelly, and Xiu (2020) conclude that a neural network model with three layers displays better out-of-sample performance than traditional or more advanced return forecasting benchmarks. Chen, Pelger, and Zhu (2023) estimate the stochastic discount factor using a deeper neural network applied to all U.S. equity data combined with a substantial set of macroeconomic and firm-specific information.

Several authors recently have used machine learning to predict mutual fund performance. Li and Rossi (2021) use machine learning methods to predict mutual fund returns according to the stocks they hold. Kaniel et al. (2023) show that fund momentum and fund flow are the most important predictors of future risk-adjusted fund performance, while stock characteristics are not predictive. DeMiguel et al. (2023) document that machine learning based solely on fund characteristics identifies managers whose skill is not sufficiently offset by diseconomies of scale.

Machine learning has shown not only the power to predict stock returns but also a prospect that it can select better-performing mutual funds. Can actively managed mutual funds still provide value to Main Street investors in the era of AI? We try to answer this question, and present evidence that mutual fund managers still demonstrate human wisdom beyond the power of AI.

Mapping historical stock information and market context to future stock performance is theoretically the best way to use ANN in stock selection, so that approach is applied in most

studies in the literature. Yet such mapping is highly subject to noise for various reasons. For example, the timing of the realization of stock returns, given a specific combination of features and context, is most likely unstable and dependent on many factors. Even when investors can identify stocks with perfect fundamental features in the most favorable market condition, when superior excess returns will be realized, if it eventually happens, is pretty much random. Such stocks might even have rather poor returns in some future periods. Thus, the practical applicability of such ANN models is questionable.

Avramov, Cheng, and Metzker (2023) empirically document that machine learning-based investments are profitable primarily for difficult-to-arbitrage stocks and during severe limits-to-arbitrage market states. They show that machine learning signals also prompt a remarkably high turnover and often require taking rather extreme long-short positions. Neither feature is desirable from a practical investment management perspective.

To avoid these problems, we focus prediction on stocks that are most likely to be traded by actively managed U.S. open-end equity mutual funds. Chen, Jegadeesh, and Wermers (2000) show that stocks held by funds do not outperform other stocks, while stocks purchased by funds have significantly higher returns than stocks they sell. Active stock trades represent a stronger manager opinion about a stock than the passive decision to hold a current position in the stock, because holding may be driven by non-performance-related reasons such as concerns over transactions costs or alignment with mandated fund style.[1] Chen, Jegadeesh, and Wermers (2000) conclude that the consensus opinion of the entire mutual fund industry about a stock provides superior information about the value of that stock.

In our study of the use of machine learning, we focus predictions on the stocks that actively managed U.S. open-end equity mutual funds are most likely to purchase. In other words, we aim to predict the most important investment decision that fund managers make. Further, we identify stocks that mutual funds trade in but whose trades are not predicted by AI, and show the impressive performance of these stocks. More important, we evaluate the out-of-sample performance of *individual* mutual funds that hold different amounts of these stocks in their portfolios. In other words, we present evidence about human wisdom in the era of AI investing.

It has been well documented that mutual funds on average do not outperform their benchmarks, because expense ratios are not trivial (Grinblatt and Titman, 1989, 1993; Carhart,

[1] The U. S. Securities and Exchange Commission now requires mutual funds to maintain a minimum of 80% of the value of the portfolio in securities that are consistent with the fund's advertised style (see Investment Company Names, Investment Company Act Release No. IC-24828. 01-1967.pdf (govinfo.gov)).

1997). Yet studies have uncovered findings that do favor active mutual fund performance. Glode's (2011) theoretical model finds consistent empirical evidence showing that active mutual fund performance can be underestimated because of performance measure misspecification, which would result in a finding of underperformance after costs. Cremers and Petajisto (2009) develop a measure of active management that can predict fund performance. They document that funds with high active share significantly and persistently outperform their benchmarks. Similarly, Petajisto (2013) investigates the types and degrees of active management of mutual funds and finds that the most active stock pickers can beat benchmarks even after expenses.

Several authors have shown that fund performance cannot be attributable solely to luck, and that fund managers exercise skill. Kosowski et al. (2006) use several bootstrap approaches to show strong evidence of persistent superior performance of fund managers rather than luck. Cuthbertson, Nitzsche, and O'Sullivan (2008) find similar evidence suggesting the existence of stock picking ability among top-performing mutual funds in the U.K. Berk and van Binsbergen (2015) also show that fund managers have skills, and differences in skills persist for as long as ten years. These findings confirm the importance of fund manager skills.

Given that fund managers have skills that can lead to superior performance, it is expected that an investment strategy be relatively stable, as it is unlikely that fund managers change their perceived "winning formulas" from day to day. Thus, if ANN can learn such winning formulas, performance would be expected to be more stable and reliable. Our study follows this approach to examine how ANN can learn from the trading decisions of fund managers and whether such a strategy is profitable. More important, we shed light on an understanding of whether mutual fund managers have human wisdom beyond AI power.

To examine machine learning on the mutual fund trade decision making process, we first ask whether an ANN can predict the stocks that would experience the most positive changes in mutual fund holdings depending on stock characteristics and market information. This would shed light on understanding how managers make decisions on stock trades, including trade preferences and the impact of important variables on trading decisions.

Second, to examine the value of the trading insights that ANN can learn from fund managers, we further investigate whether stock trades predicted by ANN can deliver better performance than a number of benchmarks. We compare the performance of the investment strategy based on ANN prediction with the performance of a strategy that simply mimics the last major trades of fund managers, that invests in mutual fund portfolio stocks, and that invests in all universe stocks in the market.

We find that the ANN can predict mutual fund trades at an accuracy of 62.1 percent in an out-of-sample test. The ANN can also identify a number of preferences of fund managers in their stock purchases, such as herding, past performance (short-term reversal, momentum), percentage change in shares outstanding, return on invested capital, and others.

Our results suggest that fund managers have the ability to select good stocks and outperform the market. More important, a portfolio of stocks selected by fund managers but not selected by ANN shows superior performance. This indicates that fund managers have skills that are unlearnable by ANN. Further, we show that funds with higher levels of non-AI skills tend to have higher returns. More important, investors in general can use machine learning to identify and invest in these mutual funds.

Our results have important implications for practitioners beyond contributing to the literature on mutual fund performance and fund manager skill, as well as the application of machine learning in finance. First, the stock selection approach based on ANN learning from investment decisions of fund managers is a valid approach. In fact, good returns generated by the ANN strategy firmly confirm our conjecture that machines can learn from the "winning formulas" of fund managers. There are potential benefits for investors to follow a similar approach and incorporate further advanced features of neural networks.

Second, we show evidence that confirms the unlearnable skill of fund managers. The most successful strategy is to invest in stocks in which mutual funds are increasing their purchase, unforeseen by an ANN model. This demonstrates that fund managers exercise insights in their trading, which validates their value in the mutual fund industry. Finally, we shed light on how to identify fund managers with valuable non-AI skills in practice. On average, net-of-expense returns of equity funds with high non-AI skill level significantly outperform those of low non-AI skill by about 3.72% annually in terms of Carhart four-factor alpha.

The paper is organized as follows. Section 2 introduces the ANN model employed to learn from fund managers, while Section 3 describes the data. Section 4 presents empirical results on the predictability and investment performance of ANN and the importance of non-AI skills of fund managers. Section 5 concludes.

2. Methodology

2.1. Artificial Neural Network (ANN) model

We construct a feed-forward ANN as illustrated in Figure 1. The ANN has multiple layers classified into three types: input layer, hidden layers, and output layer. Within each layer, there

are multiple processing units called nodes or neurons. The input layer has N_X nodes corresponding to N_X input variables of the model. In other words, each node in the input layer receives the value of an input variable of the model.

The hidden layers may include one or multiple layers that are connected via one-directional linkages, where every node in a layer receives the information from all nodes of the previous layer and passes the processed information on to the nodes of the next layer. The first hidden layer receives information from the input layer, and the last hidden layer passes its information on to the output layer.

The output layer may have one or multiple nodes representing one or multiple outputs of the network, which are obtained from processing input from the last hidden layer.

Except for the nodes of the input layer, each node in any layer processes the information it receives using an *activation function* $f(\cdot)$. We denote the input of node j in layer l as $z_j^{[l]}$, and the output of that node is $a_j^{[l]} = f^{[l]}\left(z_j^{[l]}\right)$, where $f^{[l]}(\cdot)$ is the activation function of all nodes in layer l. The input $z_j^{[l]}$ of layer l is a linear combination of the outputs of the nodes in layer $l-1$. Specifically:

$$z_j^{[l]} = b_j^{[l]} + \sum_{i=1}^{N_{l-1}} w_{ji}^{[l]} \times a_i^{[l-1]} \quad (1)$$

where $b_j^{[l]}$is the constant term of the linear combination, $w_{ji}^{[l]}$is the parameter associated with the input from node i of layer $l-1$ to node j of layer l, and N_{l-1} is the number of nodes in layer $l-1$.

We denote $W_j^{[l]}$as the $1 \times N_{l-1}$ vector that includes all input parameters of node j in layer l, and $W^{[l]}$ as the $N_l \times N_{l-1}$ matrix that includes all input parameters of all nodes in layer l. The information flow structure of the whole neural network can be presented as follows:

$$a^{[l]} = f^{[l]}\left(z^{[l]}\right) \text{ for } l = 1, 2, \cdots, L \quad (2)$$

$$z^{[l]} = b^{[l]} + W^{[l]} \times a^{[l-1]} \text{ for } l = 1, 2, \cdots, L \quad (3)$$

where $a^{[l]}$ is the $N_l \times 1$ vector that includes all outputs of layer l, $b^{[l]}$ is the $N_l \times 1$ vector of constant terms of all nodes in layer l, $z^{[l]}$ is the $N_l \times 1$ vector of the final inputs of the activation function of the nodes in layer l, and the activation function $f^{[l]}(\cdot)$ is applied to all elements of the input variable in layer l.

For the input layer ($l = 1$), the activation function takes the values of the input variables and produces the same outputs. This means that if we denote X as the $N_X \times 1$ input vector, the output of this layer is $a^{[1]} = X$. We follow the common approach in the literature (see, for

example, Gu, Kelly, and Xiu, 2020; Chen, Pelger, and Zhu, 2024; and Kaniel et al., 2023) to use the rectified linear activation unit (ReLU) function as the activation function for all hidden layers, which, for any input variable u, is defined as:

$$ReLU(u) = \max(0, u) \tag{4}$$

For the output layer, we use the $logistic$ function, also known as the $sigmoid$ function, as the activation function. This serves the purpose of the ANN in our study to classify stocks into binary outputs of *Good* and *No Good* categories, denoted as *G* and *NG*, respectively. We investigate only the long-only trading strategies based on ANN where good stocks (*G*) are purchased. Long-short strategies are not suitable in mutual fund industry data, because the mutual fund industry as a whole rarely takes short positions in stocks.[2]

The *sigmoid* function will produce the output of range $[0,1]$, which represents the probability a stock will be classified in the *G* class. Specifically, for any input variable u, the $sigmoid$ function is defined as:

$$sigmoid(u) = \frac{1}{1+e^{-u}} \tag{5}$$

The parameters $W^{[l]}$ and constant terms $b^{[l]}$ of the ANN are estimated so as to minimize the cost function of the network, which is calculated from the difference between the estimated outputs of the network and the actual value of the outputs. We follow Gu, Kelly, and Xiu (2020) and use the adaptive moment estimation algorithm (ADAM) as the solver for this optimization problem.

A reasonably complex ANN might provide a good fit for any non-linear relation between input and output variables, which will prompt scepticism about overfitting. A complex ANN can perfectly fit an input-output relation by simply remembering every single observation of the dataset, but it is likely to perform poorly on new data.

To account for this potential issue, we implement several regularization practices commonly used in the literature. First, we divide the sample into three separate parts, including training sample, validation sample, and testing sample. The ANN is trained (i.e., estimated) using the training sample with various values of the hyperparameters such as learning rate and regularization term. Then, each estimated ANN is evaluated using the validation sample to identify the best setting for the hyperparameters. As the validation sample is not known to the networks when they are trained, this hyperparameter tuning practice is essentially choosing the values of the hyperparameters that deliver the best out-of-sample performance.

[2] In our full sample period of 2003-2022 with 381,327 firm-quarter observations, there are only 303 instances of net negative positions in a stock and only 134 stocks are shorted.

The performance of the selected ANN is finally back-tested using the test sample, which is unknown to the ANN during the hyperparameter tuning process. We implement an expanding window approach, in which we expand the training sample to use the maximum amount of data available, and take the two years after the training sample as the validation sample. The estimated ANN will then be used to classify stocks to invest in the next quarter, which is the testing sample. We will examine both the predictability and investment performance of the ANN strategy using this testing sample.

To be specific, the ANN is first trained using the available data of 2003Q3-2010Q4, validated using 2011Q1-2012Q4 data and tested using 2013Q1 data. The next training sample, validation sample and testing sample are 2003Q3-2011Q1, 2011Q2-2013Q1, and 2013Q2, respectively.

To further mitigate the potential overfitting problem, we implement a simple network structure for the ANN that includes only three layers: one input layer, one hidden layer, and one output layer. The three-layer ANN is among the most popular structure in the literature of neural network applications in finance. The number of nodes in the hidden layer normally varies, ranging from three (Eakins and Stansell, 2003), five (Ticknor, 2013), ten (Wang et al., 2011), twenty (Liao and Wang, 2010), to 100 neurons (Qiu, Song, and Akagi, 2016). To find a lower bound on the performance of machine learning methods, we use a parsimonious structure with three neurons in the hidden layer. In robustness checks, we re-examine results using various alternative structures of the number of neurons in the hidden layer, as well as more hidden layers overall.

All networks we investigate have the same input and output layers. In the input layer the number of neurons equals the number of input variables of the network, and in the output layer one neuron produces the scalar output of the probability of a stock is classified as *G*.

Finally, we incorporate other ways in the estimation of our model to tackle over-fitting including l_1 penalization, early stopping, and ensemble following recommendations in Gu, Kelly, and Xiu (2020, 2021), and Chen, Pelger, and Zhu (2024), among others.[3] We follow Gu, Kelly, and Xiu (2020, 2021) to use the l_1 penalty term ranging from 10^{-5} to 10^{-3} and ensemble neural networks estimated at ten different random seeds. We also set the dropout probability at

[3] l_1 penalization includes a penalty for the absolute value of the estimated model parameters (multiplied by a penalty term) in the loss function of the optimization, which has the effect of pushing the unimportant parameter values toward zeros. Early stopping is used to cease the model estimation early if the loss function does not show significant improvement after several iterations. This helps to prevent the optimizer from overfitting trivial patterns in the dataset. The ensemble method uses the average of the fitted values from a number of fitted models to obtain the final prediction.

0.05 (see Chen, Pelger, and Zhu, 2024). We estimate the ANN model using TensorFlow in Python 3.9.

2.2. Stock classification using mutual fund holding data

We use the ANN to classify all stocks on the NYSE, Nasdaq, and the AMEX into *Good* (*G*) or *No Good* (*NG*) class. Under the assumption that mutual fund managers have insights on identifying good stocks to buy, those that on average earn superior risk-adjusted returns, we conjecture that stocks that mutual funds aggressively purchase are likely to be such superior stocks. More specifically, we use changes in the fractional holding of a stock by all U.S. domestic actively managed open-end equity funds, denoted as *Trades*[4], to classify a stock.

Following Chen, Jegadeesh, and Wermers (2000), we calculate the fractional holding of a stock at any quarter as the number of shares of the stock held by all funds at the end of the quarter divided by the total number of shares outstanding of the stock at that time. At the end of a quarter, we classify stocks that have positive *Trades* as *G*, and the remaining as *NG*. As the investment strategy based on ANN prediction in our study is long-only, we do not bet against stocks with zero or negative *Trades*.

In order to predict if a stock belongs to the *G* or *NG* class, we provide the ANN with relevant information that may influence the stock's future performance, including the stock trading activities, fundamental information about the company, and a variety of macroeconomic variables. The ANN will be trained to use these input data to predict the classification of the stocks in the next quarter.

2.3. Investment strategy based on ANN stock classification

At the end of every quarter starting in 2012Q4, we use data available up to that date to estimate the ANN to predict the classifications of U.S. stocks in the next quarter, and invest accordingly. Since the output value of the ANN for a stock represents the probability that the stock belongs to *G* class, we classify stocks with ANN outputs higher than 0.5 as *G* class and the remaining stocks as *NG*. We create a long only strategy based on the ANN prediction in a similar spirit to Chen et al. (2000)'s trade portfolio. Particularly, the ANN strategy will buy all stocks that are

[4] Using holding changes is more justifiable than using holding level since changes reflect the updated view of fund managers on a stock. When mutual funds have accumulated a significant position in a stock, they cannot divest it in a short period of time without suffering an adverse price impact, even when managers are pessimistic about the stock's prospects. In other cases, fund managers might be uncertain about the future of a stock in which they have heavily invested and decide to hold the position. If the classification is based on holdings level, we risk treating these stocks as superior stocks while managers are not that confident about them. When managers increase their positions in an investment, they explicitly show confidence about the stock. Our use of changes in holdings to classify good stocks is consistent with findings in Chen, Jegadeesh, and Wermers (2000) that stocks held by funds do not outperform other stocks while stocks purchased by funds have significantly higher returns than stocks they sell.

classified as *G* according to the magnitude of the ANN predicted probability. The weight of a *G* stock in a quarter is calculated as its ANN output and market capitalization as below:

$$w_{i,t}^{ANN} = \frac{O_{i,t} \times P_{i,t-1} \times Shrout_{i,t-1}}{\sum_{All\ G\ stocks} O_{i,t} \times P_{i,t-1} \times Shrout_{i,t-1}} \quad (6)$$

where $O_{i,t}$ is the predicted probability outcome of the ANN for stock *i* in quarter *t* (obtained at the beginning of quarter *t*), $P_{i,t-1}$ and $Shrout_{i,t-1}$ are the price and number of shares outstanding of stock i at the end of quarter *t-1*. We obtain similar results when the stocks are weighted using market capitalization, equal weights, or predicted output without the market capitalization multiplication. The results of these alternative weighting schemes are available upon request.

We compare the investment performance of the ANN fund using several benchmarks. First, we use a simple strategy that is based only on the historical trade of fund managers in the previous quarter. We therefore construct the *Historical Trades Benchmark (HT-BM)*, a benchmark that invests in stocks with positive *Trades* in the previous quarter. In a similar spirit with the weighting scheme of the ANN, we calculate the weight of a stock *i* at the beginning of quarter *t* as:

$$w_{i,t}^{HT} = \frac{Trades_{i,t-1} \times P_{i,t-1} \times Shrout_{i,t-1}}{\sum_{All\ stocks\ with\ positive\ Trades_{i,t-1}} Trades_{i,t-1} \times P_{i,t-1} \times Shrout_{i,t-1}} \quad (7)$$

We refer to this benchmark as *Historical Trade Benchmark (HT-BM).*

Second, we compare the performance of the ANN fund and that of the mutual fund industry in general. We combine holdings of all U.S. equity funds in our sample to create a *Total Fund Benchmark (TF-BM)*, which tracks the performance of the whole mutual fund industry. The weight of a stock at the beginning of a quarter for this benchmark is calculated as the fractional holding of the stock held by the whole mutual fund industry at the end of the previous quarter multiplied with the market capitalization of the stock at the end of the previous quarter, and divided by the sum of these values of all firms invested in by the whole mutual fund industry.

Finally, we compare the ANN fund performance to the whole stock universe. We consider the market capitalization weighted average and the naïve equal-weighted average of all stocks available in the sample at any portfolio formation date. These benchmarks are referred to as the *Market-Capitalization-Weight Benchmark (VW-BM)* and the *Equal-Weight Benchmark (EW-BM)*, respectively.

3. Data

3.1. Mutual fund holding and performance

Our stock data is obtained from the Center for Research in Security Prices (CRSP) stock return files and the CRSP Survivor-Bias-Free Mutual Fund database.[5] We consider only U.S. domestic actively managed open-end equity funds (OEFs).[6]

The sample period starts in January 2003 and ends in December 2022. Mutual fund families introduced different share classes in the 1990s. Because different share classes have the same holdings composition, we aggregate all the observations pertaining to different share classes into one observation. To describe the qualitative attributes of funds (e.g., objectives and year of origination), we use the observation of the oldest fund. For total net assets (TNA) under management, we sum the TNA of the different share classes. For the quantitative attributes of funds (e.g., returns and expenses), we take the weighted average of the attributes of the individual share classes, where the weights are the lagged TNA of the individual share classes.

To address the incubation bias documented by Elton, Gruber, and Blake (2001) and Evans (2010), we follow the procedure proposed by Kacperczyk, Siam, and Zheng (2008) to exclude observations prior to the reported fund-starting year and observations where names of funds are missing from the CRSP database. In addition, we include newly established funds only after they first reach at least US$5 million in assets under management. Once they reach the first threshold of $5 million, they stay in the sample until the end.[7]

3.2. ANN input data

At the beginning of any quarter, for every stock in the U.S. market, we prepare 105 inputs that the ANN will use to classify it into either *G* class or *NG* class. First, we include all 64 firm-level variables proposed in Freyberger, Neuhierl, and Weber (2020), many of them are well-

[5] In 2010, CRSP switched its holdings source from Thomson-Reuters and Lipper to Lipper's Global Holdings Feed alone. There are data irregularities in CRSP mutual fund holdings, particularly prior to 2010. Stocks with a change in CUSIPs in CRSP stock files are commonly duplicated, causing errors in portfolio holdings. We confirm these errors by examining actual holdings disclosures available on the SEC EDGAR website, and correct the data.

[6] Domestic equity funds have "E" and "D" in the first two characters of the CRSP Style Code (variable: *crsp_obj_cd*), where CRSP maps the objective codes of Strategic Insights, Wiesenberger, and Lipper into a continuous series. The third character "S" in the variable—*crsp_obj_cd*—indicates a sector fund. Some mutual funds switch between sector funds and non-sector funds. We identify portfolios of non-sector funds using their style codes at the beginning of each calendar quarter. We use the CRSP variable *index_fund_flag* to separate actively managed funds from passively managed funds. Prior to 2008, we follow Shive and Yun (2013) to classify funds as index funds.

[7] To mitigate the incubation bias, Kacperczyk, Sialm, and Zheng (2008) and Cici, Gibson, and Moussawi (2010) exclude funds with assets of less than $5 million in the previous month. This filter might unintentionally exclude Ameritor Security Trust (crsp_fundno: 005371; Ticker: ASTRX) from the calculations in January-May 1996 because its TNA was below $5 million in the previous months, although it was a seasoned fund. Its inception date was December 1939, and its initial TNA $49 million was first recorded in December 1961.

documented in the literature to influence asset prices. To account for possible lags in the release of a company's financial statements, we take only data from statements dated at least three months before the time of examination.[8]

Second, we include two important firm-level variables in stock selection, which are herding and the correlation of a stock with other stocks currently held in the mutual fund portfolio. We follow Lakonishok, Shleifer, and Vishny (1992) to define herding of stock i at the end of quarter t as:

$$Herd_{i,t} = \left|\frac{B_{i,t}}{B_{i,t}+S_{i,t}} - P_t\right| - E\left[\left|\frac{B_{i,t}}{B_{i,t}+S_{i,t}} - P_t\right|\right] \quad (8)$$

where $B_{i,t}$ is the number of funds that have positive *Trades* in stock i in quarter t; $S_{i,t}$ is the number of funds that have negative *Trades* in stock i in quarter t; P_t is the ratio of the number of positive *Trades* across all stocks and funds over the number of all non-zero *Trades* in quarter t; and $E[\cdot]$ is the expected value operator, calculated based on the fact that $B_{i,t}$ is the number of success in $(B_{i,t} + S_{i,t})$ trials of a binomial distribution with probability of success P_t. For the correlation measure, we calculate the correlation of a stock held (not held) by the total mutual fund with the value-weighted returns of the other stocks (all stocks) held by the total mutual fund using daily return data during the quarter. Table 1 provides a list of stock characteristics used as inputs for our ANN.

Third, because the future performance of each company will be significantly dependent on macroeconomic conditions, we include macroeconomic variables in the Fred_QD database from the Federal Reserve Bank of St. Louis. This includes 246 quarterly variables classified into 14 major groups, ranging from National Income and Product Accounts to Non-Household Balance Sheets. As macroeconomic variables are highly correlated, we follow the recommendation of Ludvigson and Ng (2007) to use Principal Component Analysis (PCA) to obtain a smaller group of principal component factors as the input for the ANN model.

That is at the end of a quarter, we use all historical observations available from 1993Q1 (i.e., ten years before our first fund holding data) up to the examination date (accounting for lags in data availability) to implement the PCA for each of the 14 major groups, and use the first two components of each group as the ANN input. We also follow all the variable transformation recommendations in McCracken and Ng (2020) in processing the macroeconomic data. Finally, we include industry identifier for a firm according to ten industry

[8] We adjust for seasonality of each item in the financial statements of the company before calculating the variables. Specifically, items on the balance sheet are the average values of the corresponding items over the last year, and items on the income statement are the sum of the values of the corresponding items over the last year.

classification system from Kenneth French's website, the lagged value of the *Trades* classification, and the lagged value of the total mutual fund holding of the stock.[9] Since our mutual fund holding data start in 2003Q1, we lose one observation for the calculation of *Trades*, and another observation for the lagged *Trades*. Thus, the first full set of inputs for ANN in our investigation is from the beginning of 2003Q3.

To eliminate look-ahead bias, when we predict the stock classification to invest in quarter $t+1$, we use trading related variables (for example, Beta, volatilities, maximum returns, turnovers, among others) obtained at the end of quarter t, and accounting-related and macroeconomic variables obtained at the end of quarter $t-1$.

Overall, for each stock, there are 105 input variables that the ANN uses to predict the classification in the next quarter.[10] Given enough observations, the ability to cope with a huge number of input variables is an advantage of neural networks, which could work in practical problems with millions of input variables, such as object detection and self-driving cars. In fact, for classification tasks, Baum and Haussler (1989) propose a minimum number of training observations that is at least equal to the number of parameters to be estimated by the model divided by the targeted fraction of samples incorrectly classified.

With 105 inputs and a three-layer ANN with three neurons in the hidden layer, our standard ANN model has 322 parameters to estimate (including constant terms). Even in the first training sample from 2003 through 2010, there are 80,704 firm-quarter observations, more than enough for a classification problem with a target error rate as low as 0.4 percent. The use of regularization in the ANN also helps to address potential overfitting because of the large number of model parameters.

4. Empirical results

4.1. Summary statistics

Table 2 shows the summary statistics of stock characteristics used as inputs for the ANN for the sample period from 2003Q3 through 2022Q4. For each input, we report the 25 percent quantile, median, 75 percent quantile, mean, and standard deviation of all stocks in the sample as well as stocks with positive *Trades*.

[9] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[10] To remove the impact of outlier, we winsorize our independent variables using 1 percent threshold. Any value of each independent variable that is lower than 1 percent quantile or higher than 99 percent quintile of its distribution is replaced by 1 percent quantile and 99 percent quantile correspondingly.

A comparison of all stocks and those with positive *Trades* reveals interesting findings. For example, Returns on Invested Capital (*ROIC*) of the positive *Trades* stock is significantly higher than returns of the whole stock market. This suggests that fund managers have a preference for firms with high ROIC. A similar observation can be made for lagged Value of Mutual Fund Holdings (*VMFH*), suggesting that fund managers tend to have persistent trading ideas about companies over multiple quarters.

We can also see that fund managers tend to avoid stocks with very low Earnings to Price (*E2P)*, Earnings per Share (*EPS*), Pretax Profit Margin (*PM_pretax*), Return on Assets (*ROA*), Return on Equity (*ROE*), and size (*LME*), as the 25th percentiles of these characteristics in positive *Trades* are significantly higher than the corresponding values of all stock sample. In terms of herding, it is not surprising that stocks with positive *Trades* generally have more herding than other stocks in the market, indicating the well-known herding behavior of fund managers.

4.2. ANN prediction performance evaluation

Figure 2 shows the out-of-sample prediction performance of the ANN using the "confusion matrix" and evaluation metrics commonly used in the data science literature, including *Accuracy*, *Precision* and *Recall.* Specifically,

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

where TP, True Positive, is defined as the number of actual *G* stocks (i.e., stocks that will have positive *Trades* in the following quarter) that the ANN correctly classifies; FN, False Negative, is the number of actual *G* stocks that the ANN classifies as *NG*; and FP, False Positive, is the number of actual *NG* stocks (i.e., stocks that will have non-positive *Trades* in the following quarter) that the ANN classifies as *G*.

Accuracy measures the overall predictability of the model, representing the proportion of the correct predictions among all predictions by the ANN. *Precision* captures the quality of the *G* stock prediction of the ANN, and $Recall$ reflects the ability of the ANN to select *G* stocks. There is a trade-off between *Precision* and *Recall*, as an aggressive ANN that classifies many stocks as *G* tends to deliver high $Recall$ at the cost of low *Precision*, and vice versa. We also report in Figure 2 the correlation between the percentile of the actual *Trades* value of a stock within the distribution of all *Trades* each quarter, and the corresponding output values of

the ANN that captures the likelihood that the ANN will classify a stock as *G*. We report this correlation for all stocks as well as for all *G* stocks in the out-of-sample 2013Q1 – 2022Q4 period.

The confusion matrix and evaluation metrics shown in Figure 2 confirm the ability of ANN to learn from fund managers. On average, ANN has 62.1% *Accuracy.* The *Precision* result suggests that 53.5 percent of *G* stocks that the ANN selects are correct selection. This is considerably higher than the success rate of randomly selecting *G* stocks, represented by the proportion of *G* stocks (36,539 cases of *TP* and *FN* out of the complete test sample of 94,357 stocks, or 38.7 percent). The ANN is rather conservative in choosing *G* stocks, however. On average, only 11,128 stocks, or 11.8 percent of the whole sample, are classified as *G* stocks (*FP* and *TP).* The *Recall* is also only 16.3 percent. We see positive correlation between a stock' *Trades* value with its predicted output from the ANNs. Thus, the ANN performs as expected, classifying stocks with higher *Trades* as *G* stocks at a higher level of confidence.[11]

Figure 3 shows the values of the prediction evaluation metrics calculated for each quarterly testing sample. This illustrates how the predictive performance of the ANN changes over time. While *Accuracy* is highly stable, *Precision* fluctuates around 60% throughout the sample period, with a peak of 100% at the end of 2020 and 87.5% at the end of 2019Q3. Meanwhile, *Recall* varies significantly over time, experiencing multiple highs and lows.

We conjecture that the performance of the ANN can be influenced by the volatility of the market in the training, validation, and testing sample. This can be partially observed through the plotting of the VIX value over the same sample period, which moves in a reverse pattern to that of *Recall*. To investigate this impact, we regress the prediction evaluation metrics on the average volatility of the market (measured by the average value of the VIX at the end of each quarter during the period) in the training period, validation period, and testing period.

One might argue that it is the relative uncertainty between the estimation period and the testing period that drives the out-of-sample performance of the ANN, rather than the absolute level of uncertainty. Therefore, we also use the relative volatility measure of the testing period. We first sort quarterly VIX over the combined period of training and validation into quintiles, and then assign the quarterly VIX in the testing period a quintile according to the quintile cutoffs from the combined period. The result of this investigation is presented in Table 3.

[11] Our sample consists of 38.7% *G* stocks and 61.3% *NG* stocks. Thus, a random guess will have the *Precision* of 38.7% and *Recall* of 50%. In fact, an ANN can have a *Recall* as high as 100% by predicting all stocks as *G*, which comes with the cost of high number of *NG* stocks incorrectly selected as *G* stocks. We prefer the quality of identifying the true *G* stocks (high *Precision*) to the quantity of *G* stocks selected (high *Recall*).

While the market volatility shows insignificant influence on the ANN's precision, we observe statistically significant and positive impact of the volatility during the validation period on *Recall*. It is harder to find *G* stocks in the validation sample over the two years following the training sample. When volatility in the validation period is higher, the chosen hyperparameters need to allow for more "easy going" ANN. As a result, the higher the volatility of the market during the validation period, the greater the ability of the ANN to identify *G* stocks. Meanwhile, relative volatility of the testing period is shown to have a significant negative effect on *Recall*. This is understandable, as the ANN appears to be less able to select *G* stocks when the market shows higher uncertainty than the period it is trained and validated on.

We further investigate the marginal impact of stock characteristics on the probability that they are selected as *G* stocks by ANN. We follow Gu, Kelly, and Xiu (2020) to measure the marginal impact of an input by varying its value across the deciles of its distribution while keeping all other inputs at their median and observe the change in the probability that a stock will be selected by the ANN (i.e., the output of the ANN model). We perform this calculation for every quarter of the test sample and examine the average values across the quarters. Figure 4 presents the top ten characteristics that increase, and the top ten characteristics that reduce, the probability that a stock will be selected by the ANN.

We observe that herding (*Herd*) has the strongest impact on the probability that stocks will be classified as *G*, which is consistent with Grinblatt, Titman, and Wermers (1995). Other top characteristics that enhance the chance of a stock is selected by the ANN are mostly related to stock past performance, such as short-term reversal (*r2_1*), momentum (*r12_2*), percentage change in the shares outstanding (*pΔSo)*, lagged value of the dependent variable (*VGSC*), return on invested capital (*ROIC*), and closeness to the 52-week high (*P2High*). Our findings of mutual fund preference for past performance are in line with Grinblatt, Titman, and Wermers (1995) and Chen, Jegadeesh, and Wermers (2000). Interestingly, the average correlation (*Corr*) between a stock and other stocks held by the fund positively affects its selection probability.

Characteristics that significantly reduce the probability a stock will be selected are likely size- and value- related, including bid-ask spread (*Spread*), industry-adjusted size (*LME_adj*), idiosyncratic volatility (*IdioVol*), total assets over market capitalization (*A2ME*), and book-to-market (*BTM*).[12]

[12] To further understand the importance of an input variable on stocks in different components of the confusion matrix, we carry out a similar investigation where the decile and median thresholds in each quarter are obtained

An important advantage of ANN is its ability to capture the interaction effects between variables. Figure 5 illustrates the pairwise interaction effects between Herding (the most important input variables in the previous investigation) and a handful of other variables. We follow Gu, Kelly, and Xiu (2020) to capture the interaction effect between two variables by varying a variable across the deciles of its distribution while keeping the other variable (*Herd* in this case) at one of the five specific levels, which are the minimum value, 25% percentile, median (50% percentile), 75% percentile, and maximum value of its distribution. All other variables are kept at their corresponding median values. The interaction effect is recognisable if there are differences in the impact curve (the curve showing how different decile values of a variable affect the selection probability) when the controlling variable *Herd* varies. If all of the impact curves of a variable are parallel, which means they only shift vertically along with the changes in *Herd*, this indicates no interaction between the examining variable and *Herd*. To clearly present the difference between the shapes of these curves for each variable, in the second row of Figure 5, we shift all of these curves to zero starting point.

It can be seen clearly that while there is little interaction between *Unexp_volm* and *Herd*, there are noticeable interaction effects between *Herd* and the other variables. Particularly, the impacts of correlation (*corr*), short-term reversal (*r2_1*), Price-to-cost margin (*PCM*), and Price to 52-week high price (*P2High*) become much weaker for stocks with high *Herd* (light blue line). We also observe multiple twisting points between the impact curves of *corr* and between those of *P2High*, which further confirm that, for stocks with different *Herd* levels, these variables affect the selection probability differently.

4.3. ANN portfolio performance

Beyond evaluating the ANN's ability to select *G* stocks, which are stocks that fund managers consider good, we are interested in the performance of portfolios constructed using ANN. We consider three levels of transaction cost, from 0 basis point (frictionless market), 10 basis points, to a conservative 50 basis points per trade. We report in Table 4, for each strategy and with each trading cost assumption, the total holding-period return from the beginning of 2013Q1 through the end of 2022Q4, the average quarterly returns net of transaction cost, the volatility measured by the standard deviation of quarterly returns, the tail risk measured by 95 percent Value-at-Risk (VaR) of the quarterly returns, the conventional Sharpe ratio calculated as the ratio of average quarterly return in excess of three-month T-bill over volatility, and the

from True Positive stocks, False Positive stocks, False Negative stocks, and True Negative stocks separately. We observe that the majority of important inputs shown in the investigation of all stocks stay important in the selection probability of stocks in the components of the confusion matrix. The results are available upon request.

Sharpe_VaR ratio calculated as the average quarterly return in excess of three-month T-bill over the value of VaR. Thus, the conventional Sharpe ratio represents the reward to a unit of volatility, while the Sharpe_VaR ratio represents the reward to a unit of tail risk.

The results presented in Table 4 suggest the ability of fund managers to add value in the investment process. The total return, average return and tail risk-adjusted returns of the *Total Fund Benchmark (TF-BM)* are higher than those of *VW-BM* and *EW-BM*. This evidence of fund managers' skills is consistent with the findings of Cuthbertson, Nitzsche, and O'Sullivan (2008) and Berk and van Binsbergen (2015). However, fund managers' skills are not easy to follow. A strategy following historical trades of fund managers *(HT-BM)* performs worse than the market capitalization weighted index. These observations establish the need for more sophisticated methods to extract information from mutual fund trades to develop desirable trading strategies. As an attempt to achieve this, our ANN delivers desirable good performance. As mentioned above, the ANN fund outperforms all benchmarks in terms of total return, average return, and tail risk adjusted return. More important, its performance is consistent, whether the market is frictionless or there are transaction costs. A disadvantage is that the ANN strategy has higher volatility than all other benchmarks, and higher tail risk than most of the benchmarks. This leads to a relatively weak volatility-based Sharpe ratio of the ANN strategy. Nevertheless, the tail-risk based Sharpe ratio of the ANN is still considerably higher than those of all benchmarks at all transaction cost levels.

We also observe that transaction cost has an important impact on the performance of the ANN strategy and *HT-BM*. When transaction costs increase from 0 to 50bps, the average returns of these portfolios reduce by about 0.8 percentage points (from 0.049 to 0.041) and 0.7 percentage points (from 0.03 to 0.023), respectively. Meanwhile, the Sharpe_Volatility ratios drop by about 17 percent, from 0.326 to 0.269 for ANN fund and 25 percent from 0.316 to 0.237 for *HT-BM*. Similarly, Sharpe_VaR ratios of ANN fund and *HT-BM* also decrease by about 21 and 29 percent, respectively. Transaction cost has only minimal impact on the other strategies. The *VW-BM* experiences the least reduction in average returns and Sharpe ratios.

To illustrate how these strategies perform in different periods of the market, Figure 6 presents the overtime cumulative performance of the ANN strategy and the benchmarks in frictionless market (0 basis point cost) and high friction market (50 basis points cost). The ANN strategy consistently stays on top of other strategies in most of the examining period when there is no transaction cost. It also exhibits significantly better returns than the benchmarks in multiple periods, such as in 2013, 2018, 2020, and 2021. However, in an aggressive transaction

cost environment, the ANN strategy needs to rely on these return jumps to catch up and stay on top of the other strategies.

4.4. Can human investing still have value?

We now investigate in detail the classification of stocks according to the ANN. The intersection between the classification of the ANN and the actual classification of stocks creates four groups of stocks: TP (stocks that are chosen by both the ANN and fund managers), FP (stocks that are chosen by the ANN and not by fund managers), FN (stocks that are chosen by fund managers and not by the ANN), and TN (stocks that are not chosen by either the ANN or fund managers). In quarterly testing samples, we create confusion matrix portfolios consisting of stocks in these groups, and refer to them as PortTP, PortFP, PortFN, and PortTN, respectively.

Table 5 presents the median values of the characteristics of the stocks in these portfolios over the test samples. This table reveals interesting patterns in the way ANN and fund managers select good stocks to invest. Consistent with discussion so far, PortFN, which is not selected by the ANN, includes stocks that are considerably larger and traded actively (as shown in high values of *TA*, *LME*, and *Turn*) while the ANN, PortFP, appears to favor stocks with higher operating leverage (*OpLev*), and higher past returns (*r12_2*). While ANN prefers stocks with greater momentum, skillful fund managers tend to be contrarians (as shown in low *r12-2*, *r12-7*, *r2-1*, *r36-13*, *r6-2*, and *P2High*). This suggests that skillful fund managers tend to have their own opinions rather than following the herd.

Furthermore, the non-AI skill of fund managers tends to identify undervalued stocks of companies that might be experiencing difficulty. PortFN has values of a number of fundamental ratios at the unfavorable end, including low profitability (*CT*, *Prof*, *RNOA*, *ROIC*, *S2TA*), low investment (*Invest*, *pΔBE*, *ΔINVT*), high uncertainty (*Std_volm*), and low unexplained volume (*Unexp_volm*). Although these measures are not as low as those of PortTN, they are poorer than the corresponding values in PortTP and PortFP.

Our finding about the contrarian tendency of PortFN is consistent with Wei, Wermers and Yao (2015). They document that stocks mostly selected by contrarian funds have fewer investment activities and higher uncertainty. More important, they show that contrarian fund managers deliver superior performance, and their strategies not only differ from herds but also show low commonality among themselves. This is confirmed by our observation of the low *Herd* and *Corr* values of PortFN as compared to those of PortFP. Their finding that contrarian managers have distinctive strategies and do not "mechanically" trade together against herd also supports our observation of non-AI skill of fund managers.

Yet stocks selected by skillful fund managers also exhibit certain positive signals. For example, companies in PortFN exhibit high gross profit margins (*PCM*) but considerably low industry-adjusted operating profit margins (*PM_adj*). Thus, skillful fund managers show a tendency to find firms in industries with profitable products that may be facing issues in operation cost structure.

Table 6 presents the investment performance measures of the confusion matrix portfolios along with the benchmarks for the test sample. The results reveal important findings regarding the skills of fund managers as well as the quality of ANN selection. There is strong evidence that fund managers have desirable skills that are unlearnable by ANN. PortFN, which holds stocks that fund managers select but the ANN does not, has impressive performance in almost all metrics, with the highest total return and Sharpe ratios. It also has the lowest tail risk and very low volatility. In fact, PortFN outperforms the second-best strategy, PortTP, by a wide margin in risk adjusted returns. The Sharpe_Volatility and Sharpe_VaR ratios of PortFN are substantially higher than those of PortTP across all transaction costs.

Figure 7 shows the proportion of common stocks held over two consecutive quarters of the confusion matrix portfolios during the testing period. PortFN has relatively low turnover as compared to the other confusion matrix portfolios. On average, the composition retention rate of PortFN fluctuates in the range of 12 percent to 50 percent most of the time, with an average of 36.9 percent. This is higher than the average value of PortTP (25.8 percent) and significantly higher than that of PortFP (2.3 percent). As a result, the performance of PortFN is not significantly affected by transaction costs. It continues to be the best portfolio in all transaction cost settings, and there remains a considerable difference between the performance of PortFN and those of other portfolios. A focus on firms with good quality in accounting metrics such as ROC and ROIC as well as momentum might explain why the ANN driven PortFP has a steady and high portfolio turnover.

Second, the results indicate that the ANN can learn a certain part of the fund managers' skills to deliver decent stock selection decision. PortTP is the second-best performing strategy. In fact, the total returns and average returns of PortTP are close to those of PortFN. PortTP has high volatility and tail risk, leading to much lower Sharpe ratios than PortFN. It is the risk measure that worsens the risk-adjusted performance of PortTP. Also, among stocks that are not selected by fund managers, those chosen by ANN perform better than those that are not. As expected, PortTN is the poorest performing strategy among all portfolios and benchmarks examined. It earns even lower return and lower risk-adjusted return than the worst benchmark of simple equally weighted portfolio of all stocks in the market.

Panel B of Table 6 shows the DeMiguel et al. (2009) *t*-statistics for the difference between the Sharpe ratios of the strategies. These confirm the superior performance of PortFN. The Sharpe ratios of PortTP are also significantly higher than those of PortFP, PortTN, and EW-BM. PortTP also performs better than HT-BM and VW-BM, although the difference is not statistically significant.

4.5. Non-AI skills of fund managers

Results so far suggest that fund managers have desirable skills in active management that are not learnable by Artificial Intelligence. We further investigate this phenomenon by examining the relation between the non-AI skill level of fund managers and their actual performance. We investigate this via a portfolio sorting analysis. At the beginning of every quarter from 2013Q2 to 2022Q4, we sort mutual funds according to the proportion of stocks with positive *Trades* in their portfolio not selected by the ANN in the previous quarter before the sorting time. We then divide the mutual fund universe into deciles based on their non-AI skills at the beginning of each quarter and calculate returns of these quintiles in the next quarter. Decile 1 are funds with the lowest non-AI skill level while Decile 10 are funds with the highest non-AI skill level.

Table 7 presents the returns of the Decile portfolios and the difference between the returns of Decile 10 (high non-AI skill) and Decile 1 (low non-AI skill). We also report the alphas and the estimated coefficients of Carhart (1997) four-factor regressions of the excess returns of the Decile portfolios, and of a long-short strategy that buys Decile 10 and sells Decile 1. Panel A reports results of Gross return and Panel B presents results of Reported Net-of-expense Returns.

The results in Table 7 strongly confirm the value of fund manager non-AI skills. The average return of Decile 10 is significantly higher than that of Decile 1 in term of both gross return and reported net-of-expense return, as shown in significant positive alphas of D10-D1. Specifically, the value-weighted net-of-expense returns of equity funds with high non-AI skill level are significantly higher than returns of low non-AI skill by about 0.93 percent per quarter (or 3.72 percent annually) in terms of Carhart four-factor alpha. Moreover, results in Table 7 also reveal that funds with low non-AI skills (Decile 1) deliver significantly negative alpha. Meanwhile, fund managers with high non-AI skills (Decile 10) earn significantly positive alpha in their gross returns but insignificant alpha in their reported net-of-expense returns. This could indicate that they extract all rents by charging higher expense ratio. To our knowledge, this is the first evidence in the literature that confirms the value of non-AI skills of fund managers, as well as the practical benefits of taking advantage of such skills. In the risk exposure analysis,

funds with higher non-AI skills tend to maintain a full exposure to the market and load more on SMB.[13]

To further investigate the persistence of non-AI skill levels of fund managers, we present in Figure 8 the transition matrix of non-AI skill deciles, showing the probability that a fund moves from one decile to another in the next quarter, averaged over the period from 2013 to 2022. If fund managers' non-AI skill were not persistent, we would expect the probability of a fund moving between deciles to be around 10 percent. Yet as Figure 8 shows, the prominent diagonal strongly suggests persistent non-AI skill of fund managers. A fund will always be more likely to stay in its current decile over quarters, with a probability significantly higher than 10 percent. Note that persistence is especially strong for very low and very high non-AI skill, as 18 percent of funds in Decile 1 continue to stay in Decile 1 in the next quarter, and the corresponding number for Decile 10 is 19.5 percent. Consider that if fund managers have very low (high) non-AI skill levels, it is highly unlikely that they will acquire (lose) such skills in such a short time. This is in line with evidence of the persistence of fund manager skill in Berk and van Binsbergen (2015). Our findings further confirm that it is practical that public investors can identify and invest in fund managers with high non-AI skills in a profitable and consistent manner.

4.6. Robustness checks

4.6.1. Different machine learning models

Our base model specifies a simple structure of the ANN with only one hidden layer including three neurons. Next, we investigate the use of different machine learning models, including Support Vector Machine (SVM) and alternative ANN structures with different complexity.

We first investigate the classification problem using a completely different machine learning model. That is, we use a Support Vector Machine (SVM) to select stocks following exactly the same data processing pipeline as for the ANN. In fact, while the power of deep neural networks might be more suitable to very large datasets with millions of observations, many studies in the literature document the ability of SVM to outperform neural networks in medium or small datasets (Byvatov et al., 2003; Behzad et al., 2009; Dong et al., 2022). At the beginning of every quarter starting from 2013Q1, we estimate the SVM using the training data from 2003Q3 with different hyperparameter settings, and then use the last two years of data as

[13] To address that the public might not know mutual fund holdings data right at the portfolio disclosure date, in a robustness check we allow a one-quarter lag to define the non-AI skill level of a fund manager. For example, at the beginning of 2013Q3, public investors can obtain the ANN selections and mutual fund holdings at the end of 2013Q1 as well as the corresponding non-AI skill level of funds in 2013Q1. We obtain similar results where the long-short strategy (D10-D1) earns significantly positive Carhart (1997) alpha.

a validation sample to choose the best hyperparameters. The chosen estimated SVM will be used to classify stocks in the coming quarter.[14]

Figure 9 reports the out-of-sample confusion matrix and prediction performance evaluation metrics of the SVM model. The results confirm our findings above in that SVM has very similar predictive performance to ANN. Thus, only a certain part of fund manager skills can be learned by a machine learning model. Machine learning models can identify only 21.32 percent of stocks selected by fund managers, and the *Precision* of their selection is around 53.4 percent, nearly 15 percentage point higher than a naïve random choice (which is 38.7 percent as discussed earlier).

In examining the performance of ANN models with alternative complexity, we vary the numbers of layers and neurons in each layer in the ANN model. Figure 10 to 12 present the out-of-sample confusion matrix of the ANN with one hidden layer that has only 2 neurons (referred to as ANN_2), a larger model with one hidden layer that has 32 neurons (referred to as ANN_32), and a deeper model with 4 layers, in which the numbers of neurons in layer 1 to 4 are 64, 32, 16, and 8, respectively (referred to as ANN_64_32_16_8)[15]. The predictability performances of these models are similar to each other, and are similar to the base ANN model, in terms of both *Accuracy, Precision,* and *Recall*. These findings suggest that there is a certain part of the mutual fund managers' skills that is relatively easy to be learned and that can be captured by even a simple machine learning model, as shown in the *Precision* of over 50% in all models.

Tables 8-11 report the performance of the investment strategies based on SVM, ANN_2, ANN_32, and ANN_64_32_16_8. The results confirm the value of what machine learnings can gain in learning from fund managers. As in the base ANN model, the machine learning models perform especially well in total returns and average returns. In fact, they are the best-performing strategies in terms of returns before adjusting for risk and continue to outperform other benchmarks in tail risk-adjusted returns. Their performance, however, is worsen in volatility-adjusted returns because of significantly higher volatility than other benchmarks.

[14] We use the Radial Basis Function (RBF) Kernel, which is the most generalized form of kernelization. In tuning hyperparameters for SVM, we set the regularization parameter *c* to range from 1 (default value in scikit-learn package in Python) to 0.25, and the kernel coefficient gamma to range from 0.0001 to 0.1. We also use the default value of gamma in scikit-learn of "scale" as well as the "auto" value in the package (which is equal to 1 divided by the number of input variables in the model).

[15] We follow the common convention in setting up deep neural networks to have the number of neurons in each layer a multiplier of 8, and the number of neurons in each layer reduces gradually from the first hidden layer to the last hidden layer.

Their performance is also significantly affected by trading costs due to their high turnover. The complex model of ANN_64_32_16_8 experiences the strongest impact. Both returns and adjusted returns of this strategy decrease in the aggressive cost setting (50bps), as compared to the frictionless setting (0bp). Specifically, Sharpe-Volatility and Sharpe-VaR both drops by over 5 percentage points. This is not surprising, as a more complex model can "remember" more from the training data and be more sensitive to input variables. As a result, the model changes selected stocks more frequently than more parsimonious models.

Using these alternative models, we also investigate the value of AI and non-AI skills of fund managers and examine the performance of portfolios constructed according to the confusion matrices. The results shown in Table 8-11 are consistent with our main findings, suggesting that the non-AI skills of fund managers are highly valuable. PortFN is the best-performing portfolio in terms of risk-adjusted returns at all transaction cost settings. We continue to observe that PortTP is the second-best portfolio in risk-adjusted returns in almost all models. As discussed earlier, the most complex model, ANN_64_32_16_8, performs worse than other models. In this setting, PortTP only performs better than PortTN and EW-BM, while being poorer than other portfolios and benchmarks. This, along with our analysis above, suggests that complex ANN models tend not to be suitable for fund manager skill learning. This might suggest that our dataset is not large enough to make complex ANN model effective.

Finally, PortTN is still the worst portfolio of all. It consistently underperforms all other benchmarks in terms of both returns and risk-adjusted return. This confirms that in general fund managers know how to avoid poorly performing stocks and that skill can be picked up quite easily by machine learning.

4.6.2. The value of non-AI skills using different models

Panel A-D of Table 12 report the results using the alternative models (SVM, ANN_2, ANN_32, and ANN_64_32_16_8, respectively) in examining the value of non-AI skills. The findings consistently show a significant difference between the future returns of the top non-AI funds and the bottom ones. Thus, while the AI skills of fund managers might be easily picked up by machine learning, their non-AI skills are very difficult for the machine to acquire. In addition, funds with high non-AI skills tend to maintain a full exposure to the market and load more on SMB than funds with low non-AI skills. These skills are the most valuable asset of fund managers.

More important, these results confirm the value of machine learning for Main Street investors. It is straightforward for investors to use machine learning to identify good fund

managers. Investors can use various machine learning models ranging from simple or complicated specifications to identify high non-AI skill managers and invest in them.

4.6.3. Filtering insignificant positive trades

In our standard framework, we classify any stock that has positive *Trades* as *G* stock. Thus, there are cases where stocks with very small positive *Trades* are considered *G*, which might add noise to the ANN estimation. To address this possible concern, we re-estimate our base ANN model in which only stocks that are in top 90 percent of *Trades* in a period are classified as *G* stocks. In other words, we consider stocks at the bottom 10 percent of *Trades* as noise and classify them as *NG* stocks.

The result of this investigation (unreported here to preserve space and is available upon request) suggests that using the filtered positive *Trades* slightly reduce the predictability of the ANN, with an increase in *Accuracy* and a reduction in *Precision* and *Recall*. This is mainly because the data in which the ANN learns from becomes more unbalanced (a smaller proportion of stocks are classified as *G* stocks). Thus, the ANN has stronger tendency to predict a stock to be *NG*, increasing *Accuracy* overall but reducing the other measures. Nevertheless, we still observe a strong performance of the ANN fund as compared to the benchmarks, the outstanding performance of PortFN, and the prominent difference in returns between funds with high non-AI level and those with low non-AI level, yielding significant Carhart (1997) alpha. Therefore, our results are robust with respect to the exclusion of stocks with very small *Trades* as *G* stocks. However, we opt for the inclusion of such stocks since they are likely to be mega-capitalization stocks and a modest change in fractional ownership is still informative.

4.7. Can AI learn from funds with high non-AI level?

It is well known that the majority of mutual funds underperform S&P 500 Index. The volume of these underperforming funds could generate considerable noises which could hinder the AI learning. We have documented the superior performance of funds with high non-AI level and the persistence of these funds. Next, we examine whether the machines can enhance their performance by learning directly from these persistent top-performing funds, rather than from a pool of all mutual funds. This will shed light on an essential question whether the non-AI skills of funds is learnable when AI directly focuses on them.

To carry out this investigation, we first obtain the list of top 50 percent of funds according to the non-AI level (as measured in the previous sections) of every quarter. At the end of every quarter from 2012Q4 to 2022Q3, we use the estimated ANN at that time to obtain the top non-AI funds for the following quarter. In order to have a reasonable amount of data to train the ANN, the list of top non-AI funds for quarters in the 2003-2012 periods is obtained

based on the stock classifications using the first available ANN (estimated at the end of 2012Q4) and the input data available at the beginning of each quarter. We then consider the mutual fund industry to include only these top funds and re-estimate the ANN similarly to the standard framework. We refer to this as ANN_TopNonAI. The approach we use to obtain top non-AI funds in 2003-2012 introduces certain look ahead bias.

Figure 13 reports the ability of ANN_TopNonAI in identifying *G* and *NG* stocks. It is clear that the ANN's predictability performance is weaker than the base framework, especially for *Precision*, which is 44.58% as compared to 53.49%. Thus, it appears that it is harder for the ANN to learn from the top non-AI funds. Furthermore, the performance of the strategy based on ANN_TopNonAI is also rather weak, as reported in Table 13 and 14. Specifically, in Table 13, the ANN Fund is now the worst performing strategy, with risk and return measures even poorer than EW-BM. This is true even before taking into account transaction costs. In Table 14, there is also no significant difference between the returns of top non-AI funds and bottom non-AI funds when the non-AI levels are obtained based on this new ANN model. In this investigation, we classify all mutual funds in the industry according to the new non-AI levels estimated using the ANN_TopNonAI. The poor performance of the portfolio sorting practice suggests that non-AI deciles now include funds with mixed quality. For example, top non-AI deciles now includes both superior and inferior funds, and the same applies to bottom non-AI deciles. In other words, the fund skills picked up by the ANN are not valuable anymore, and therefore funds with high or low level of these skills do not perform significantly different. In short, non-AI skills of fund managers are hard to learn. This conclusion is even stronger given the fact that the ANN_TopNonAI is estimated using data with potential look-ahead bias.

5. Conclusion

We investigate performance of Artificial Neural Network (ANN) and their ability to learn from fund managers in stock selection and investment strategy. We first show that ANN can predict fund managers' stock selection with 62.1 percent accuracy in out-of-sample tests. Herding and past performance are found to increase the probability that ANN will select a stock, while size and value-related firm characteristics have a negative impact on stock selection probability.

An ANN strategy provides higher investment returns than the total fund and other benchmarks, including those based on funds' historical trades and value-weighted as well as equally weighted portfolios. At the same time, our results consistently suggest that fund managers have critical skills that are non-learnable by AI. Portfolios of stocks selected by fund

managers but not selected by ANN show superior performance. Fund managers with higher levels of non-AI skills tend to have higher returns.

Our results are robust to a variety of sensitivity checks. More complicated structures of ANN and Support Vector Machine (SVM), an alternative machine learning model, yield similar results. We also obtain consistent findings when filtering out stocks with very small *Trades* among *G* stocks.

Our work confirms the ability of Artificial Neural Network to learn from fund managers but also highlights the importance of human skills in portfolio selection and investment strategy. We also find evidence that these valuable skills are very hard to learn even when the AI is trained directly on holdings of funds with high levels of such skills. The findings shed light on how to identify fund managers with valuable non-AI skills in practice, which is surely useful for investors.

References

Avramov, D., Cheng, S. and Metzker, L., 2023. Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*, 69(5), pp.2587-2619.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, pp.627-635.

Baum, E. B., and Haussler, D., 1989. What size net gives valid generalization. *Neural Computation*, 1, pp.151-160.

Behzad, M., Asghari, K., Eazi, M., and Palhang, M., 2009. Generalization performance of support vector machines and neural networks in runoff modelling. *Expert Systems with applications*, 36(4), pp.7624-7629.

Berk, J.B., and van Binsbergen, J.H., 2015. Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1), pp.1-20.

Brockett, P.L., Cooper, W.W., Golden, L.L., and Pitaktong, U., 1994. A neural network method for obtaining an early warning of insurer insolvency. *Journal of Risk and Insurance*, 61(3), pp.402-424.

Byvatov, E., Fechner, U., Sadowski, J., and Schneider, G., 2003. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences*, 43(6), pp.1882-1889.

Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance*, 52(1), pp.57-82.

Chen, H., Jegadeesh, N., and Wermers, R., 2000. The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. *Journal of Financial and Quantitative Analysis*, 35(3), pp.:343-368.

Chen, L., Pelger, M. and Zhu, J., 2024. Deep learning in asset pricing. *Management Science*, 70(2), pp:671-1342.

Cici, G., Gibson, S. and Moussawi, R., 2010. Mutual fund performance when parent firms simultaneously manage hedge funds. *Journal of Financial Intermediation*, *19*(2), pp.169-187.

Cremers, K.M., and Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies*, 22(9), pp.3329-3365.

Cuthbertson, K., Nitzsche, D. and O'Sullivan, N., 2008. UK mutual fund performance: Skill or luck? *Journal of Empirical Financ*e, 15(4), pp.613-634.

DeMiguel, V., Garlappi, L., and Uppal, R., 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5), pp.1915-1953.

DeMiguel, V., Gil-Bazo, J., Nogales, F., and Santos, A., 2023, Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, 150 (3), 103737.

Dong, Y., Zhou, S., Xing, L., Chen, Y., Ren, Z., Dong, Y., and Zhang, X., 2022. Deep learning methods may not outperform other machine learning methods on analyzing genomic studies. *Frontiers in Genetics*, 13, 992070.

Eakins, S.G., and Stansell, S.R., 2003. Can value-based stock selection criteria yield superior risk-adjusted returns: An application of neural networks. *International Review of Financial Analysis*, 12(1), pp.83-97.

Elton, E.J., Gruber, M.J., and Blake, C.R., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance*, *56*(6), pp.2415-2430.

Erel, I., Stern, L.H., Tan, C., and Weisbach, M.S., 2021. Selecting directors using machine learning. *Review of Financial Studies*, 34(7), pp.3226-3264.

Evans, R.B., 2010. Mutual fund incubation. *Journal of Finance*, *65*(4), pp.1581-1611.

Fletcher, D., and Goss, E., 1993. Forecasting with neural networks: an application using bankruptcy data. *Information & Management*, 24(3), pp.159-167.

Freyberger, J., Neuhierl, A., and Weber, M., 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5), pp.2326-2377.

Glode, V., 2011. Why mutual funds "underperform." *Journal of Financial Economics*, 99(3), pp.546-559.

Grinblatt, M., and Titman, S., 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business*, 62(3), pp.393-416.

Grinblatt, M., and Titman, S., 1993. Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business*, 66(1), pp.47-68.

Grinblatt, M., Titman, S., and Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *American Economic Review*, 85(5), pp.1088-1105.

Gu, S., Kelly, B., and Xiu, D., 2020. Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), pp.2223-2273.

Gu, S., Kelly, B., and Xiu, D., 2021. Autoencoder asset pricing models. *Journal of Econometrics*, 222(1), pp.429-450.

Guresen, E., Kayakutlu, G., and Daim, T.U., 2011. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8), pp.10389-10397.

Heaton, J. B., Polson, N. G., and Witte, J. H., 2016. Deep learning in finance. Preprint, https://arxiv.org/abs/1602.06561.

Kacperczyk, M., Sialm, C., and Zheng, L., 2008. Unobserved actions of mutual funds. *Review of Financial Studies*, 21(6), pp.2379-2416.

Kaniel, R., Lin, Z., Pelger, M., and Van Nieuwerburgh, S., 2023, Machine-learning the skill of mutual fund managers. *Journal of Financial Economics*, 150(1), pp.94-138.

Ko, P.C., and Lin, P.C., 2008. Resource allocation neural network in portfolio selection. *Expert Systems with Applications*, 35(1-2), pp.330-337.

Kosowski, R., Timmermann, A., Wermers, R., and White, H., 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, 61(6), pp.2551-2595.

Lakonishok, J., Shleifer, A., and Vishny, R.W., 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32, pp.23-44.

Li, B., and Rossi, A., 2021. Selecting mutual funds from the stocks they hold: A machine learning approach. Working Paper, Georgetown University.

Liao, Z., and Wang, J., 2010. Forecasting model of global stock index by stochastic time effective neural network. *Expert Systems with Applications*, 37(1), pp.834-841.

Lin, C.M., Huang, J.J., Gen, M., and Tzeng, G.H., 2006. Recurrent neural network for dynamic portfolio selection. *Applied Mathematics and Computation*, 175(2), pp.1139-1146.

Ludvigson, S.C., and Ng, S., 2007. The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1), pp.171-222.

McCracken, M., and Ng, S., 2020. FRED-QD: A quarterly database for macroeconomic research (No. w26872). *National Bureau of Economic Research*.

Newey, W.K., and West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), pp.703-708.

Petajisto, A., 2013. Active share and mutual fund performance. *Financial Analysts Journal*, 69(4), pp.73-93.

Qiu, M., Song, Y., and Akagi, F., 2016. Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons & Fractals*, 85, pp.1-7.

Shive, S., and Yun, H., 2013. Are mutual funds sitting ducks? *Journal of Financial Economics*, 107(1), pp.220-237.

Ticknor, J.L., 2013. A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14), pp.5501-5506.

Wang, J.Z., Wang, J.J., Zhang, Z.G., and Guo, S.P., 2011. Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, 38(11), pp.14346-14355.

Wei, K.D., Wermers, R. and Yao, T., 2015. Uncommon value: The characteristics and investment performance of contrarian funds. *Management Science*, 61(10), pp.2394-2414.

Figure 1: Artificial Neural Network structure

This figure shows the structure of a typical Artificial Neural Network. Each node in the figure represents a neuron. An arrow from a neuron *i* to another neuron *j* indicates that the output of neuron *i* is used as the input for neuron *j*.

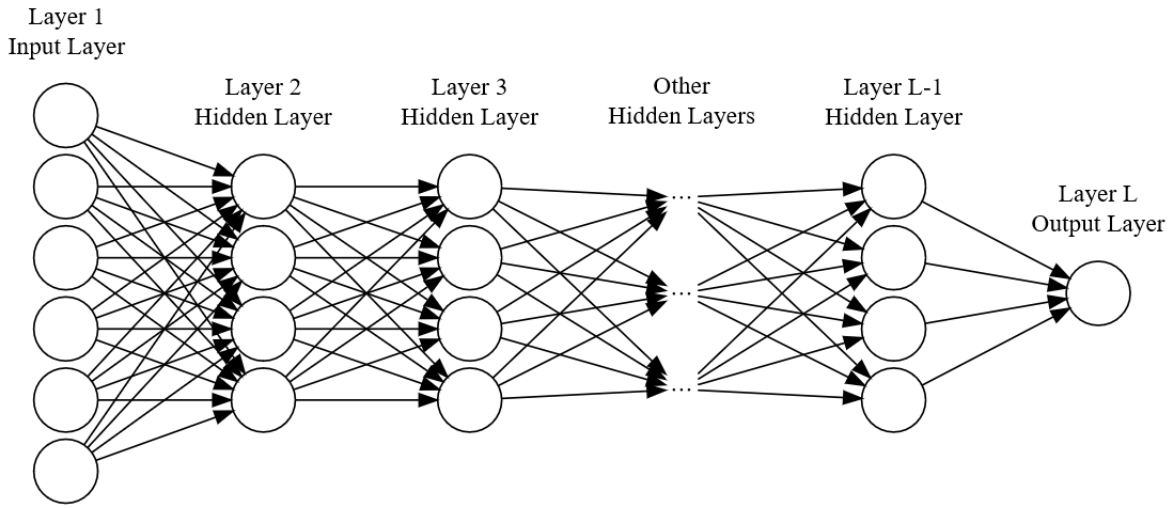
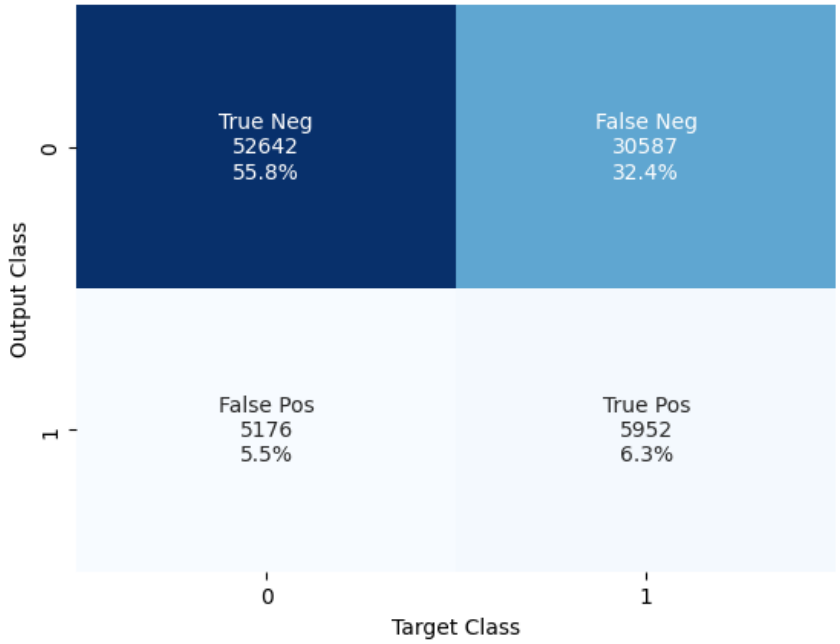


Figure 2: Out-of-sample confusion matrix of ANN

This figure shows the performance of the ANN in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the ANN and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	62.10%
Precision: TP/(TP+FP)	53.49%
Recall: TP/(TP+ FN)	16.29%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	24.45%
G stocks	20.26%

Figure 3: Out-of-sample prediction performance of ANN over time

This figure shows the values of *Accuracy*, *Precision* and *Recall* of the ANN in quarterly testing samples from 2013Q1–2022Q4. In each quarter, the ANN is estimated using data available up to the previous quarter (accounting for lags in making the data available to the public) to predict stock classification. The figure also presents the market volatility of the corresponding quarters measured by the VIX (on the secondary axis).

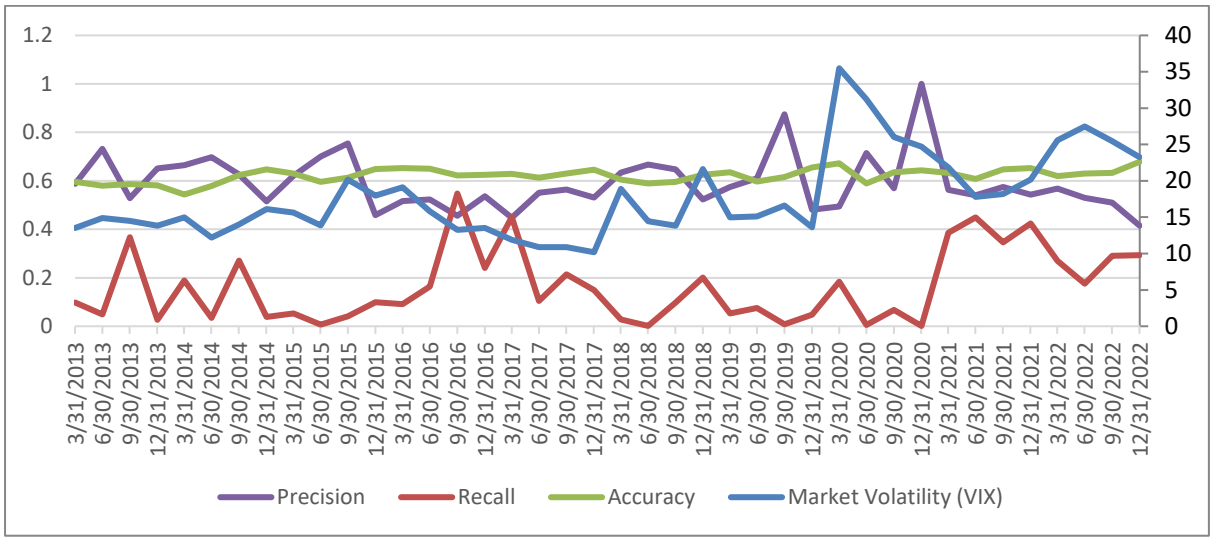
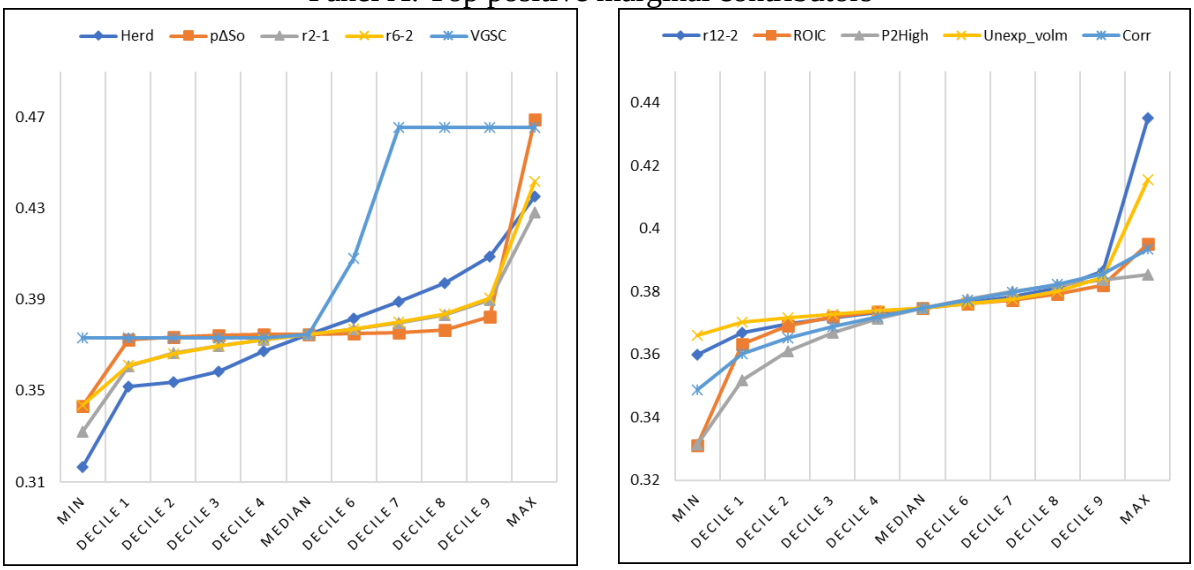


Figure 4: Ten input variables with significant marginal impact to selection probability

This figure shows the top positive and negative marginal contributors to the probability that a stock is selected by ANN. We follow Gu, Kelly, and Xiu (2020) to measure the marginal impact of an input by varying its value across the deciles of its distribution while keeping all other inputs at their median values. The marginal contribution of an input is measured as the change in the probability that a stock is classified as *G* when the input increases from its minimum to maximum value while keeping other inputs at their median values, averaged for the testing period from 2013Q1–2022Q4. In each panel, the left figure shows the top 5 impactful variables, while the right figure shows the next 5 impactful variables. Within each figure, variables whose name are on the left are more impactful.

Panel A: Top positive marginal contributors



Panel B: Top negative marginal contributors

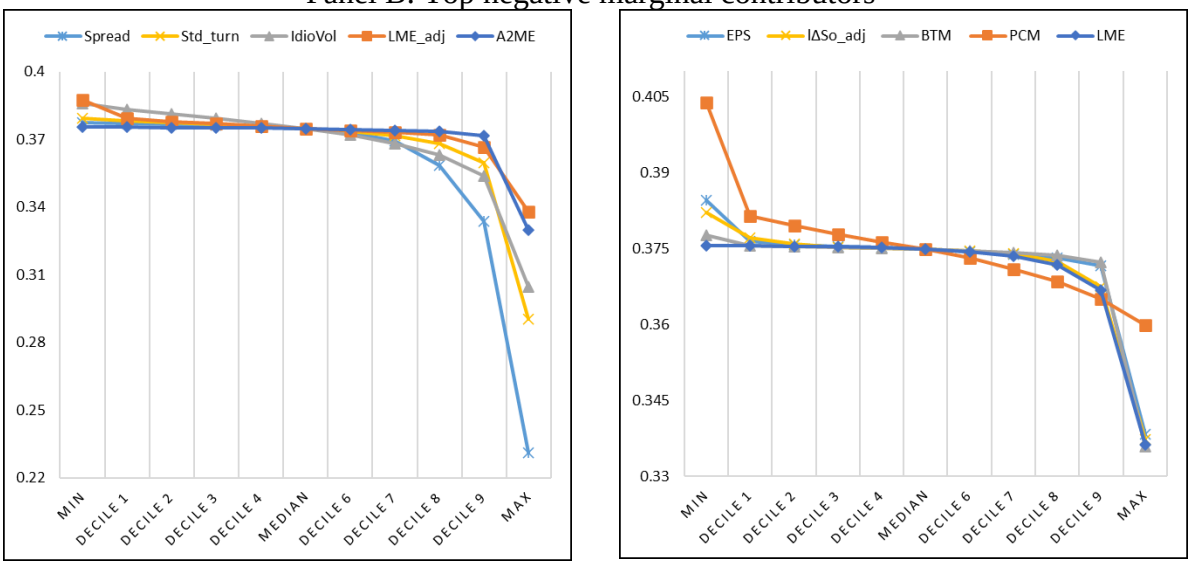


Figure 5: The impact of the interaction between herding and other variables on selection probability

The figures on the top row show the variation of the selection probability of a stock by varying a variable value across the deciles of its distribution given that *Herd* is at its minimum, 25 percent quantile, median, 75 percent quantile, or maximum levels, while keeping all other inputs at their median values. The figures on the second row rebase the corresponding figures on the top row by showing the difference in the selection probability of a stock when the examining input variable is at a decile as compared to when it is at its minimum level. The calculations are performed for every quarter in the testing period from 2013Q1–2022Q4, and the quarterly averaged results are presented.

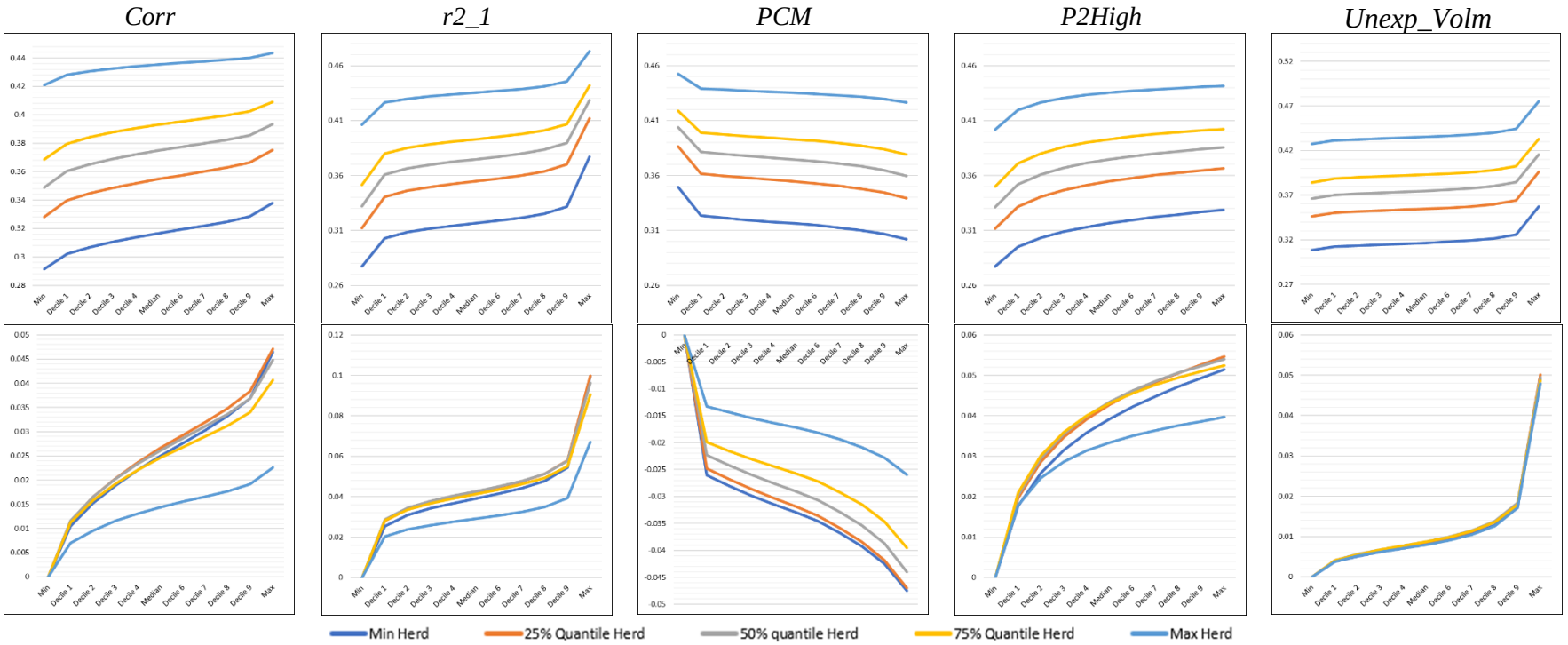
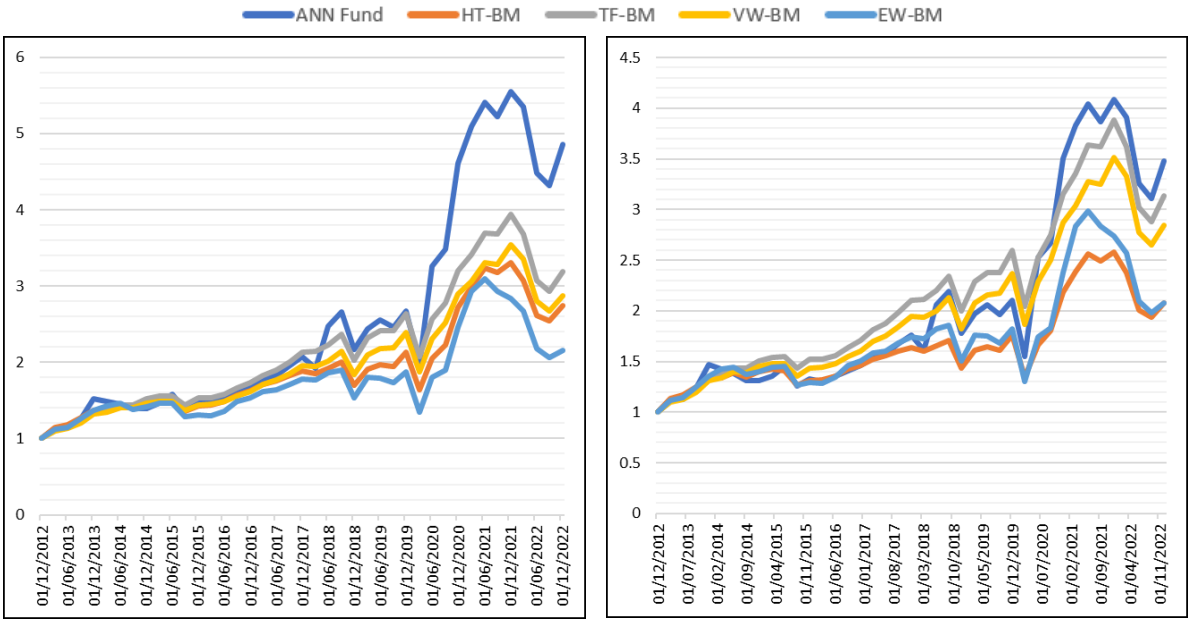


Figure 6: Performance of the ANN fund and the benchmarks over the test sample

This figure shows the cumulative performance of the strategy investing in stocks according to the ANN strategy, along with the other benchmarks, including *Historical Trades Benchmark (HT-BM)*, *Total Fund Benchmark (TF-BM)*, *Market-Capitalization-Weight Benchmark (VW-BM)*, and *Equal-Weight Benchmark (EW-BM)*. The construction of these benchmarks is described in section 2.3. The left panel shows the performance in a frictionless market with no transaction cost (0 bp). The right panel shows the performance in a market with high transaction cost of 50 basis points (bps). The performance is measured for the whole test period from the beginning of 2013Q1 to the end of 2022Q4.



0 bp transaction cost

50 bps transaction cost

Figure 7: Retention rates of portfolio composition in confusion matrix

This figure shows the percentage of common stocks held over two consecutive quarters for each confusion matrix portfolio over the test sample from 2013Q1–2022Q4. PortTP is the portfolio of stocks selected by both the ANN and the fund managers. PortFP is the portfolio of stocks selected by the ANN but not by the fund managers. PortFN is the portfolio of stocks selected by the fund managers but not by the ANN. PortTN is the portfolio of stocks not selected by either the ANN or the fund managers.

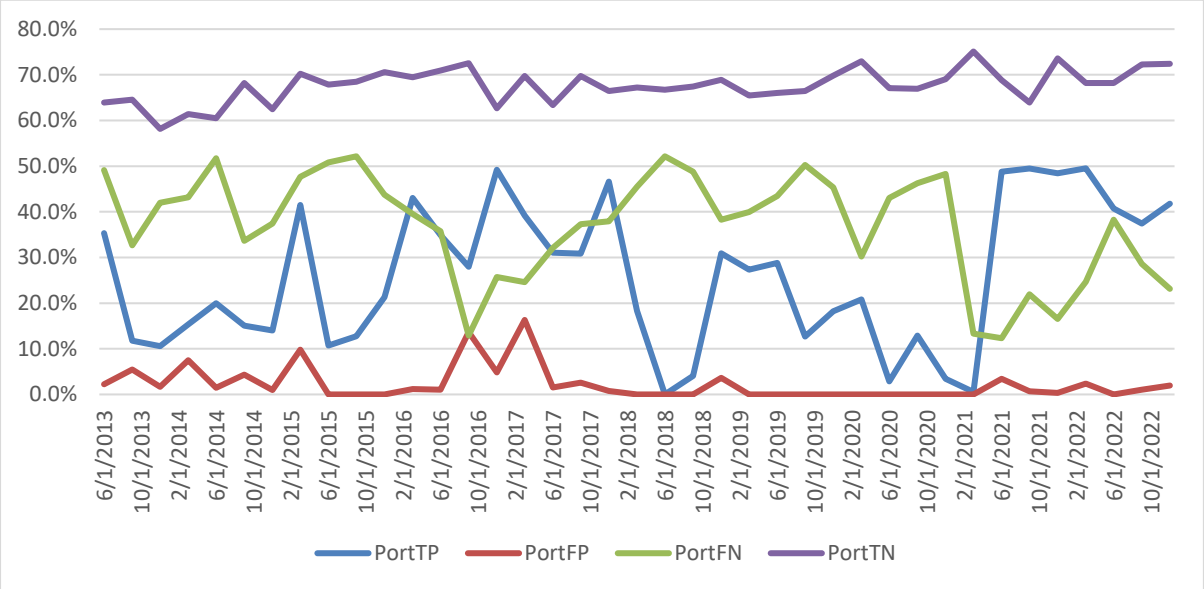


Figure 8: Transition matrix of funds in non-AI skill level quintiles

This figure shows the relative frequency with which a fund in decile *i* of the non-AI skill level decile moves into decile *j* in the next quarter, averaged over the whole sample period from 2013 through 2022. Decile 1 are funds with the lowest non-AI skill level while Decile 10 are funds with the highest non-AI skill level. The bottom row shows the t-statistics of testing the difference between the values of the diagonal of the transition matrix with 10% level.

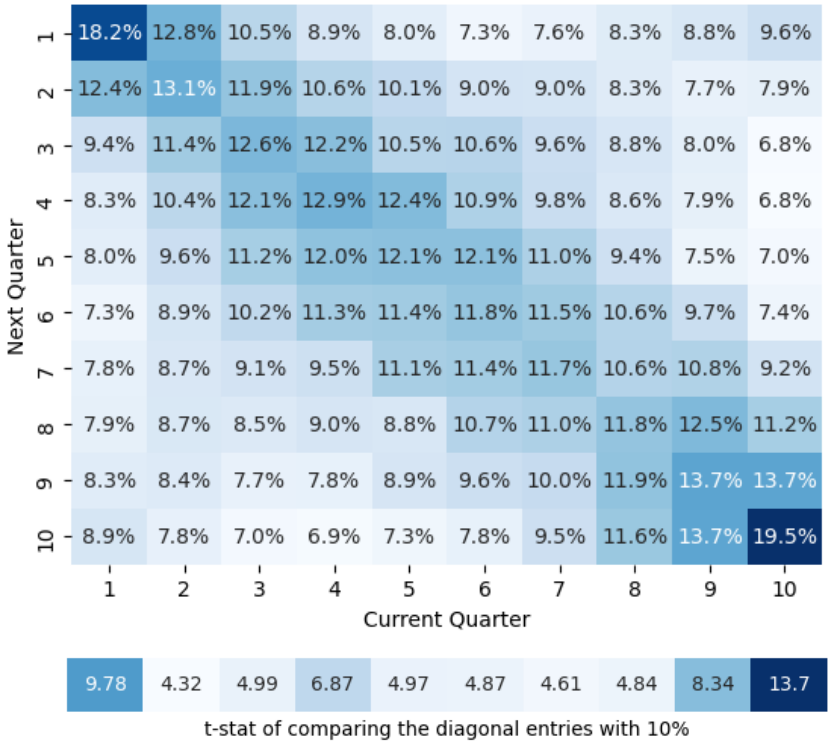
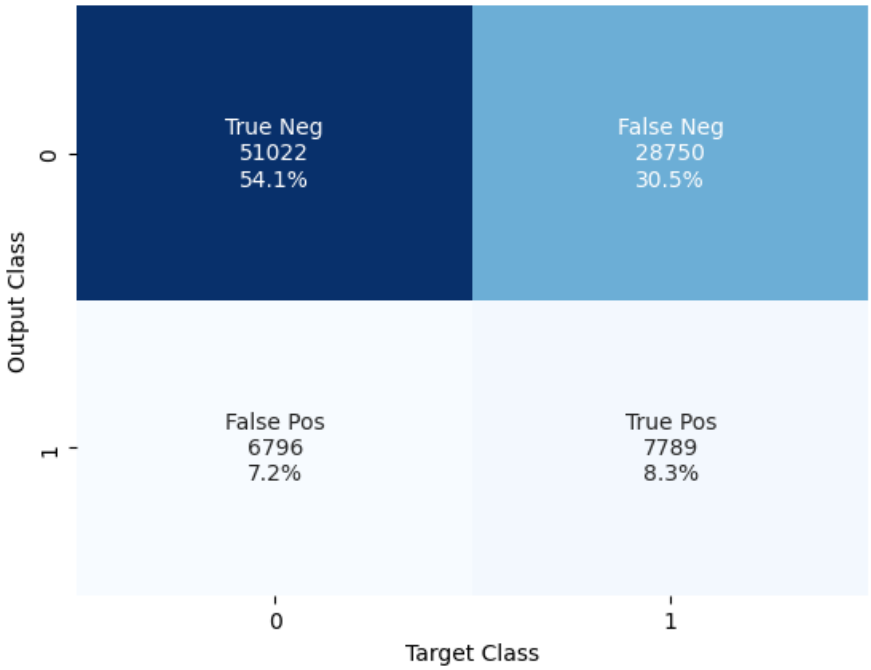


Figure 9: Out-of-sample confusion matrix of SVM

This figure shows the performance of the SVM in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the SVM and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	62.33%
Precision: TP/(TP+FP)	53.40%
Recall: TP/(TP+ FN)	21.32%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	13.70%
G stocks	7.95%

Figure 10: Out-of-sample confusion matrix of ANN_2

This figure shows the performance of the ANN_2 in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the ANN_2 and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	62.05%
Precision: TP/(TP+FP)	53.05%
Recall: TP/(TP+ FN)	17.49%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	24.02%
G stocks	18.61%

Figure 11: Out-of-sample confusion matrix of ANN_32

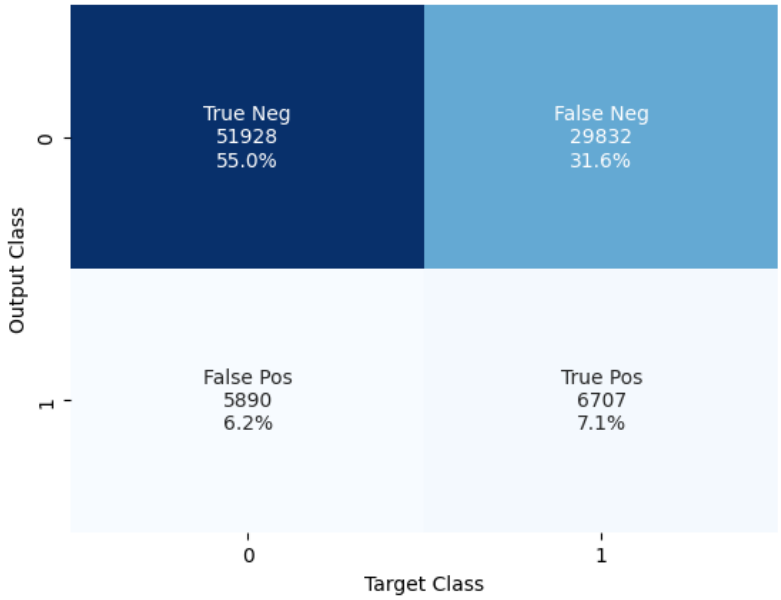
This figure shows the performance of the ANN_32 in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the ANN_32 and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	62.14%
Precision: TP/(TP+FP)	52.63%
Recall: TP/(TP+ FN)	22.25%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	24.65%
G stocks	20.95%

Figure 12: Out-of-sample confusion matrix of ANN_64_32_16_8

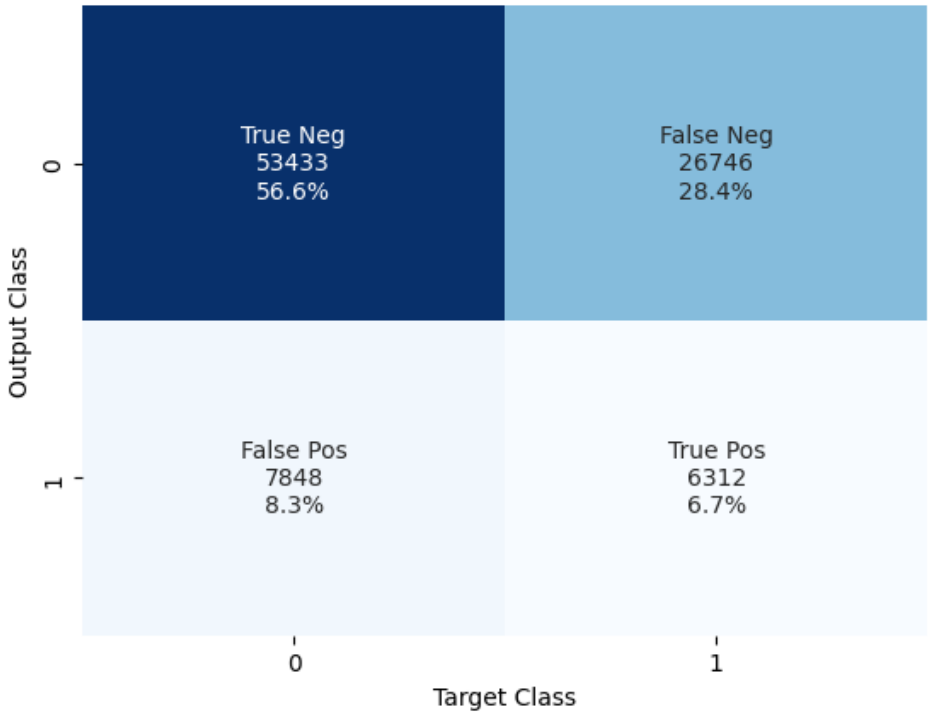
This figure shows the performance of the ANN_64_32_16_8 in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the ANN_64_32_16_8 and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	62.14%
Precision: TP/(TP+FP)	53.24%
Recall: TP/(TP+ FN)	18.36%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	24.52%
G stocks	20.84%

Figure 13: Out-of-sample confusion matrix of ANN_TopNonAI

This figure shows the performance of the ANN_TopNonAI in predicting the classes of stocks in the testing samples from 2013Q1–2022Q4. We report the number of stocks and the corresponding percentage of the number of observations in the sample falling in each category of the confusion matrix. 0 denotes *NG* stocks and 1 denotes *G* stocks. *TN* (True Negative) and *TP* (True Positive) denote stocks correctly predicted by the ANN as *NG* and *G*, respectively. *FN* (False Negative) denotes *G* stocks incorrectly predicted as *NG* while *FP* (False Positive) denotes *NG* stocks incorrectly predicted as *G*. We also report in the summary table the evaluation metrics and the correlation between the predicted classification probability of a stock estimated by the ANN_TopNonAI and the actual percentile of the stock's *Trades* among all stocks or among all *G* stocks in a quarter. A stock with a negative *Trades* value is assigned a percentile of 0.



Prediction performance	
Accuracy: (TP+TN)/(TP+FP+TN+FN)	63.33%
Precision: TP/(TP+FP)	44.58%
Recall: TP/(TP+ FN)	19.09%
Correlation: Percentile of actual *Trades* & predicted classification probability	
All stocks	20.75%
G stocks	18.67%

Table 1: List of stock characteristics

This table provides a list of stock characteristics variables. Variables are suggested by Freyberger, Neuhierl, and Weber (2020), except Herd and Corr as defined in section 3.2.

Variable	Definition
ΔINVT	Change in inventory over average total assets (TA)
ΔPI2TA	Change in PP&E and inventory over lagged TA
A2ME	Assets to market capitalization (ME)
AbOpAcc	Absolute value of operating accruals
Beta	CAPM beta
Beta_daily	CAPM beta using daily returns
BTM	Book-to-market ratio
BTM_adj	BTM - mean BTM in Fama-French 48 industry
C2D	Cash flow to total liabilities
C2TA	Cash to total assets
Corr	Correlation with total mutual fund industry holdings
CT	Capital turnover (sales to lagged total assets)
D2ME	Total debt to market capitalization (ME)
Dvd2P	Trailing 12-months dividends to price
E2P	Earnings to price (income before extraordinary items to ME)
EPS	Earnings per share
FCF	Free cash flow to book value of equity (BE)
Herd	Herding
IdioVol	Idiosyncratic volatility of Fama-French 3-factor model
Invest	% change in TA
lΔSo_adj	Log-change in split adjusted shares outstanding
Lev	Leverage
LME	Market capitalization (Price × shares outstanding) of previous month ($mil)
LME_adj	LME - mean LME in Fama-French 48 industry ($mil)
NOA	Net operating assets over lagged TA
NP2ME	Net payout to market capitalization
OP2ME	Operating payout to market capitalization
OpAcc	Operating accruals, calculated as change in non-cash working capital minus depreciation scaled by lagged total assets
OpLev	Operating leverage, calculated as (COGS + SG&A)/TA
pΔBE	% change in book value of equity (BE)
pΔGM_pΔS	% change in gross margin - % change in sales
pΔSo	% change in shares outstanding
P2High	Price to 52-week high price
PCM	Price to cost margin, calculated as (Sales - COGS)/Sales
PM	Profit margin, obtained as operating income (OI) after depreciation over sales
PM_adj	PM - mean PM in Fama-French 48 industry
PM_pretax	Pretax profit margin (pretax income over sales)
Prof	Gross profitability over BE
r_max	Maximum daily return
r12-2	Return from 12 to 2 months before prediction
r12-7	Return from 12 to 7 months before prediction
r2-1	Return from 2 to 1 months before prediction
r36-13	Return from 36 to 13 months before prediction
r6-2	Return from 6 to 2 months before prediction

Table 1 (Cont.)

Variable	Definition
RNOA	OI after depreciation to lagged net operating assets
ROA	Income before extraordinary items to lagged TA
ROC	(ME + long-term debt – TA) to cash and short-term investment
ROE	Income before extraordinary items to lagged BE
ROIC	Return on invested capital
S_g	Sales growth (%)
S2C	Sales to cash
S2OA	Sales to lagged net operating assets
S2P	Sales to price
S2TA	Sales to total assets
S2TA_adj	S2TA - mean S2TA in Fama-French 48 industry
SGA2S	Selling, general, and administrative expense (SG&A) to sales
Spread	Average daily bid-ask spread
Std_turn	Standard deviation of daily turnover
Std_volm	Standard deviation of daily volume ($mil)
TA	Total assets ($mil)
Tang	Tangibility
TobinQ	Tobin's Q
Turn	Last month's volume to shares outstanding
Turn_dt	Detrended-turnover - Market turnover
Unexp_volm	Standard unexplained volume
Vol	Standard deviation of daily return
VGSC	Value of good stock classification (1 if *Trades* is positive; 0 otherwise.)
VMFH	Value of mutual fund holdings ($mil)

Table 2: Summary Statistics

This table presents the time-series average of the mean, major quantiles, and standard deviations of the input characteristics of all stocks and stocks with positive *Trades* in our sample from the beginning of 2003Q3 through 2022Q4, respectively. A stock's trading-related variables are obtained at the end of quarter *t* (starting at 2003Q2), and accounting-related and macroeconomic variables are obtained at the end of quarter *t - 1*.

	All stocks					Stocks with positive *Trades*				
Firm characteristic	25% quantile	50% quantile	75% quantile	Mean	Standard deviation	25% quantile	50% quantile	75% quantile	Mean	Standard deviation
ΔINVT	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.01	0.01	0.03
ΔPI2TA	0.00	0.03	0.08	0.05	0.11	0.00	0.03	0.08	0.05	0.11
A2ME	0.60	1.06	1.99	2.88	7.99	0.59	1.01	1.82	2.41	6.70
AbOpAcc	0.03	0.05	0.08	0.06	0.05	0.02	0.05	0.08	0.06	0.05
Beta	0.95	1.24	1.58	1.30	0.50	0.96	1.24	1.57	1.29	0.48
Beta_daily	0.34	0.93	1.54	0.94	1.05	0.43	0.98	1.58	1.01	0.99
BTM	0.28	0.51	0.89	1.15	3.20	0.28	0.50	0.83	0.98	2.71
BTM_adj	-2.05	-0.97	-0.26	-1.18	3.80	-2.03	-0.99	-0.29	-1.32	3.46
C2D	0.03	0.15	0.30	0.12	0.58	0.06	0.16	0.32	0.18	0.51
C2TA	0.05	0.12	0.27	0.19	0.19	0.04	0.12	0.26	0.18	0.18
Corr	0.23	0.41	0.56	0.39	0.22	0.28	0.44	0.57	0.41	0.20
CT	0.56	0.93	1.45	1.12	0.82	0.61	0.97	1.50	1.17	0.82
D2ME	0.03	0.19	0.57	0.89	2.68	0.03	0.18	0.52	0.73	2.23
Dvd2P	0.00	0.00	0.01	0.01	0.02	0.00	0.00	0.02	0.01	0.02
E2P	-0.04	0.04	0.07	-0.02	0.44	-0.01	0.04	0.07	0.01	0.36
EPS	-0.23	0.63	2.22	2.04	7.93	-0.09	0.83	2.36	2.16	7.56
FCF	-0.09	0.06	0.19	0.03	0.69	-0.07	0.07	0.19	0.04	0.64
Herd	0.00	0.08	0.17	0.10	0.11	0.02	0.10	0.18	0.11	0.11
IdioVol	0.02	0.02	0.03	0.03	0.02	0.01	0.02	0.03	0.02	0.01
Invest	-3.22	5.14	15.90	9.71	26.40	-2.01	5.86	16.71	10.91	25.78
lΔSo_adj	-0.01	0.01	0.03	0.04	0.11	-0.01	0.01	0.03	0.03	0.10
Lev	0.07	0.28	0.49	0.33	0.30	0.07	0.28	0.48	0.33	0.30
LME	160.19	759.63	3111.03	5023.93	14008.99	252.63	939.86	3301.81	5061.14	13546.56
LME_adj	-4449.19	-2195.75	-253.72	662.62	13652.43	-4429.32	-2078.14	-25.54	664.43	13234.75
NOA	0.43	0.61	0.76	0.59	0.29	0.44	0.62	0.76	0.60	0.29
NP2ME	-0.01	0.00	0.01	-0.01	0.11	-0.01	0.00	0.02	-0.01	0.09
OP2ME	0.00	0.00	0.02	0.02	0.05	0.00	0.00	0.02	0.02	0.04
OpAcc	-0.07	-0.04	-0.01	-0.04	0.06	-0.07	-0.04	-0.01	-0.04	0.06
OpLev	0.46	0.78	1.25	0.95	0.71	0.47	0.79	1.25	0.97	0.71
pΔBE	-6.72	5.60	17.72	8.66	50.75	-4.73	6.59	18.41	10.38	49.20
pΔGM_pΔS	-5.40	-0.11	5.01	-2.59	51.30	-4.59	0.13	5.26	-0.93	46.12
pΔSo	-0.63	0.72	3.38	5.22	22.90	-0.74	0.76	3.62	5.77	22.21
P2High	0.64	0.80	0.90	0.75	0.19	0.68	0.82	0.91	0.78	0.17
PCM	0.25	0.39	0.57	0.40	0.25	0.25	0.39	0.57	0.41	0.24
PM	0.00	0.07	0.14	-0.05	0.69	0.02	0.08	0.15	0.01	0.54
PM_adj	0.03	0.28	4.30	5.01	12.14	0.04	0.27	4.00	4.93	12.07
PM_pretax	-0.03	0.05	0.12	-0.10	0.77	-0.01	0.06	0.13	-0.03	0.60

Table 2 (Cont.)

	All stocks					Stocks with positive *Trades*				
Firm characteristic	25% quantile	50% quantile	75% quantile	Mean	Standard deviation	25% quantile	50% quantile	75% quantile	Mean	Standard deviation
Prof	0.35	0.60	0.98	0.78	1.33	0.38	0.62	1.00	0.81	1.29
r_max	0.05	0.08	0.12	0.10	0.08	0.05	0.07	0.11	0.09	0.07
r12-2	-0.16	0.08	0.35	0.15	0.51	-0.11	0.11	0.40	0.21	0.52
r12-7	-0.13	0.04	0.24	0.09	0.36	-0.10	0.06	0.26	0.11	0.36
r2-1	-0.06	0.00	0.06	0.01	0.12	-0.05	0.01	0.06	0.01	0.11
r36-13	-0.21	0.14	0.57	0.29	0.79	-0.16	0.18	0.61	0.34	0.80
r6-2	-0.12	0.03	0.20	0.06	0.30	-0.09	0.05	0.22	0.09	0.30
RNOA	0.01	0.12	0.25	0.10	0.90	0.04	0.14	0.27	0.14	0.83
ROA	-0.03	0.04	0.09	0.01	0.17	-0.01	0.05	0.09	0.03	0.15
ROC	-2.35	1.48	7.70	4.89	28.85	-1.64	2.08	8.77	6.33	29.25
ROE	-0.06	0.08	0.18	0.03	0.53	-0.02	0.09	0.19	0.06	0.49
ROIC	0.00	0.06	0.10	0.04	0.13	0.02	0.07	0.11	0.05	0.11
S_g	-3.22	6.28	17.90	10.02	28.66	-1.92	6.96	18.45	10.99	27.02
S2C	2.77	7.33	22.21	31.19	80.68	3.01	7.95	23.72	32.69	82.77
S2OA	0.92	1.69	3.01	2.43	4.18	1.00	1.77	3.11	2.55	4.23
S2P	0.41	0.89	1.99	2.56	6.46	0.42	0.89	1.92	2.28	5.56
S2TA	0.53	0.86	1.34	1.03	0.73	0.56	0.89	1.37	1.07	0.72
S2TA_adj	-0.34	-0.05	0.31	0.05	0.64	-0.31	-0.03	0.33	0.07	0.64
SGA2S	0.13	0.24	0.41	0.38	0.58	0.12	0.23	0.39	0.33	0.46
Spread	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01
Std_turn	1.79	3.35	6.53	6.27	9.57	1.89	3.42	6.39	5.74	7.86
Std_volm	0.05	0.17	0.54	0.59	1.21	0.05	0.17	0.55	0.59	1.20
TA	194.56	818.52	3221.64	5198.43	14280.95	242.16	900.62	3347.01	5342.47	14408.13
Tang	0.38	0.49	0.58	0.49	0.16	0.38	0.49	0.58	0.48	0.16
TobinQ	0.38	0.82	1.60	1.27	1.48	0.45	0.89	1.69	1.35	1.50
Turn	2.26	4.30	7.48	6.11	6.38	2.53	4.51	7.58	6.13	5.91
Turn_dt	-0.34	0.39	1.81	1.50	4.16	-0.31	0.45	1.84	1.38	3.68
Unexp_volm	-17.76	-1.97	20.80	7.71	41.81	-16.91	-0.54	22.66	8.73	41.19
Vol	0.02	0.03	0.04	0.03	0.02	0.02	0.03	0.03	0.03	0.01
VGSC	0.00	0.06	0.99	0.41	0.49	1.00	1.00	1.00	1.00	0.00
VMFH	1.82	68.66	470.95	681.22	1777.41	12.10	113.19	535.13	698.56	1710.36

Table 3: ANN prediction evaluation and market volatility

This table shows the estimated coefficients (*t*-statistics in parentheses) of the regression of the prediction evaluation measures on market volatility (proxied by the average value of the VIX index) during the training period, validation period, and testing period, as well as the relative volatility between the testing period and the combined period of the training and validation. The dependent variables are the quarterly prediction evaluation measures over the testing period of 2013Q1–2022Q4. We first sort quarterly VIX over the combined period of training and validation into quintiles, and then assign the quarterly VIX in the testing period a quintile according to the quintile cutoffs from the combined period. *, ** and *** denote the 10%, 5% and 1% significance levels, respectively.

	Precision TP/(TP + FP)		*Recall* TP/(TP+ FN)	
Constant	0.317	0.178	0.636	0.816
	(0.563)	(0.330)	(1.027)	(1.421)
Training Vol	0.014	0.021	-0.035	-0.045*
	(0.583)	(0.875)	(-1.295)	(-1.744)
Validation Vol	-0.003	-0.004	0.021***	0.022***
	(-0.574)	(-0.655)	(3.269)	(3.522)
Prediction Vol	0.003		-0.008*	
	(0.666)		(-1.750)	
Relative Prediction Vol		0.021		-0.047***
		(1.247)		(-2.632)
Adjusted-R^2	2.54%	5.43%	30.39%	36.66%

Table 4: Performance of portfolios rebalanced according to ANN strategy and its benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN strategy or its benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly returns of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). In addition to the ANN portfolio, we also report the performance of *Historical Trades Benchmark (HT-BM)*, *Total Fund Benchmark (TF-BM)*, *Market-Capitalization-Weight Benchmark (VW-BM)*, and *Equal-Weight Benchmark (EW-BM)*. The construction of these benchmarks is described in section 2.3.

		ANN Fund	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	3.863	1.767	2.220	1.874	1.161
	Average quarterly return	0.049	0.030	0.033	0.030	0.026
	Volatility	0.147	0.088	0.083	0.081	0.113
	5% VaR	-0.172	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.326	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.277	0.184	0.200	0.183	0.128
10 bps	Total return	3.550	1.618	2.209	1.868	1.144
	Average quarterly return	0.048	0.028	0.033	0.030	0.025
	Volatility	0.146	0.088	0.083	0.081	0.113
	5% VaR	-0.174	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.314	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.265	0.173	0.199	0.182	0.127
50 bps	Total return	2.482	1.098	2.164	1.846	1.075
	Average quarterly return	0.041	0.023	0.033	0.030	0.025
	Volatility	0.145	0.088	0.083	0.081	0.113
	5% VaR	-0.178	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.269	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.219	0.132	0.196	0.181	0.122

Table 5: Confusion matrix portfolios – Median firm characteristics

For a quarterly testing sample, stock selection is predicted by the ANN estimated using data available at the end of the previous quarter. PortTP is constructed from stocks selected by both the ANN and fund managers. PortFP is constructed from stocks selected by the ANN but not by the fund managers. PortFN is constructed from stocks selected by the fund managers but not by the ANN. PortTN is constructed from stocks not selected by either the ANN or the fund managers. Each quarter, we calculate the median of firm characteristics in each of these four confusion matrix portfolios. This table shows the time-series average of quarterly median values over the testing period from 2013Q1–2022Q4.

Variables	PortTP	PortFP	PortFN	PortTN	t-statistic PortTP-PortFN
ΔINVT	0.006	0.004	0.001	0.000	1.833
ΔPI2TA	0.040	0.037	0.029	0.026	1.325
A2ME	0.916	1.007	0.999	1.101	-2.557
AbOpAcc	0.045	0.048	0.047	0.048	-0.442
Beta	1.233	1.260	1.264	1.267	-1.627
Beta_daily	1.203	1.103	0.846	0.790	9.112
BTM	0.426	0.495	0.447	0.481	-1.335
BTM_adj	-0.585	-0.773	-0.677	-0.663	2.984
C2D	0.175	0.264	0.136	0.115	2.325
C2TA	0.134	0.121	0.113	0.116	2.271
Corr	0.426	0.458	0.421	0.376	0.495
CT	1.302	1.154	0.842	0.779	4.821
D2ME	0.157	0.184	0.225	0.236	-6.057
Dvd2P	0.001	0.003	0.001	0.000	0.633
E2P	0.030	0.099	0.033	0.026	-0.333
EPS	0.820	1.339	0.898	0.477	-0.622
FCF	0.001	0.126	0.074	0.060	-0.856
Herd	0.125	0.145	0.098	0.071	3.505
IdioVol	0.021	0.019	0.020	0.022	2.189
Invest	9.831	8.419	5.135	3.910	1.672
lΔSo_adj	0.022	0.024	0.005	0.006	1.057
Lev	0.275	0.246	0.353	0.350	-4.380
LME	971.849	1256.610	1405.965	870.355	-6.442
LME_adj	-4330.459	-4503.229	-2647.889	-3063.055	-7.222
NOA	0.586	0.634	0.619	0.606	-1.357
NP2ME	0.002	0.001	0.001	0.000	1.101
OP2ME	0.004	0.004	0.003	0.001	1.384
OpAcc	-0.040	-0.043	-0.042	-0.042	0.645
OpLev	1.037	0.916	0.688	0.665	5.697
pΔBE	15.574	4.525	4.722	3.323	1.527
pΔGM_pΔS	2.330	-4.024	-0.049	-0.365	3.116
pΔSo	6.052	6.428	0.444	0.444	1.644
P2High	0.897	0.885	0.808	0.777	11.542
PCM	0.376	0.351	0.394	0.389	-1.770
PM	0.079	0.030	0.077	0.063	0.764
PM_adj	1.160	0.715	0.303	0.343	1.216
PM_pretax	0.050	0.075	0.053	0.037	-0.409

Table 5 (Cont.)

Variables	PortTP	PortFP	PortFN	PortTN	t-statistic PortTP-PortFN
Prof	0.958	0.655	0.593	0.550	1.794
r_max	0.074	0.066	0.067	0.073	3.236
r12-2	0.498	0.406	0.071	0.029	8.378
r12-7	0.184	0.192	0.044	0.020	9.097
r2-1	0.037	0.032	-0.006	-0.011	4.542
r36-13	0.239	0.221	0.191	0.109	2.352
r6-2	0.269	0.189	0.027	0.007	6.254
RNOA	0.244	0.170	0.123	0.096	1.407
ROA	0.033	0.069	0.038	0.025	-0.239
ROC	6.258	8.474	2.293	1.138	6.489
ROE	0.050	0.058	0.080	0.055	-0.416
ROIC	0.087	0.077	0.059	0.047	4.046
S_g	10.677	9.674	5.128	4.180	3.477
S2C	9.982	17.977	7.214	6.193	5.664
S2OA	2.607	2.018	1.523	1.422	1.625
S2P	0.993	0.899	0.759	0.779	4.159
S2TA	1.160	1.055	0.775	0.726	5.941
S2TA_adj	0.283	0.158	-0.043	-0.080	4.803
SGA2S	0.237	0.224	0.226	0.238	0.789
Spread	0.001	0.001	0.001	0.001	2.617
Std_turn	3.758	2.678	3.336	3.335	1.061
Std_volm	0.139	0.120	0.232	0.215	-4.678
TA	596.404	730.580	1491.348	1156.885	-14.662
Tang	0.481	0.481	0.475	0.490	1.043
TobinQ	1.429	1.464	0.868	0.744	5.344
Trades	0.011	-0.008	0.006	-0.003	3.473
Turn	5.048	3.927	4.562	4.316	1.147
Turn_dt	1.023	0.306	0.432	0.373	2.263
Unexp_volm	22.573	9.715	-2.810	-3.764	4.191
Vol	0.026	0.023	0.024	0.026	2.045
VGSC	1.000	1.000	0.313	0.000	9.401
VMFH	134.414	210.195	172.560	42.939	-3.067

Table 6: Performance of confusion matrix portfolios and the benchmarks

Panel A shows the performance measures of confusion matrix portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN strategy and its benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). Panel B shows the *t*-statistics for the hypothesis that the difference in the out-of-sample Sharpe ratio of two portfolios is zero. We follow DeMiguel et al. (2009) to calculate the *t*-statistics.

Panel A: Performance measures

		PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	4.695	1.516	5.508	0.888	1.767	2.220	1.874	1.161
	Average quarterly return	0.053	0.034	0.051	0.020	0.030	0.033	0.030	0.026
	Volatility	0.136	0.155	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.163	-0.220	-0.123	-0.166	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.376	0.207	0.585	0.214	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.313	0.146	0.404	0.107	0.184	0.200	0.183	0.128
10 bps	Total return	4.327	1.332	5.210	0.833	1.618	2.209	1.868	1.144
	Average quarterly return	0.051	0.032	0.050	0.019	0.028	0.033	0.030	0.025
	Volatility	0.135	0.155	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.165	-0.221	-0.124	-0.167	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.364	0.195	0.572	0.205	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.299	0.136	0.391	0.102	0.173	0.199	0.182	0.127
50 bps	Total return	3.074	0.719	4.146	0.629	1.098	2.164	1.846	1.075
	Average quarterly return	0.044	0.024	0.045	0.016	0.023	0.033	0.030	0.025
	Volatility	0.134	0.154	0.084	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.170	-0.227	-0.128	-0.169	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.315	0.145	0.517	0.169	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.248	0.098	0.341	0.083	0.132	0.196	0.181	0.122

Table 6 (cont.)

Panel B: Sharpe ratio statistical difference									
		PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	PortTP		3.331	-3.367	2.798	1.058	-0.197	0.321	3.235
	PortFP			-5.046	-0.293	-2.010	-2.971	-2.524	-0.165
	PortFN				7.503	6.961	5.770	6.469	6.819
	PortTN					-3.713	-7.566	-7.143	0.196
	HT-BM						-2.920	-1.648	3.978
	TF-BM							4.005	4.267
	VW-BM								3.627
	EW-BM								
		Port11	Port10	Port01	Port00	THC-BM	TF-BM	VW-BM	EW-BM
10 bps	PortTP		3.369	-3.386	2.766	1.140	-0.402	0.111	3.064
	PortFP			-5.083	-0.358	-1.977	-3.156	-2.719	-0.362
	PortFN				7.503	7.024	5.584	6.314	6.744
	PortTN					-3.532	-7.676	-7.305	0.009
	HT-BM						-3.510	-2.321	3.566
	TF-BM							3.954	4.286
	VW-BM								3.657
	EW-BM								
		Port11	Port10	Port01	Port00	THC-BM	TF-BM	VW-BM	EW-BM
50 bps	PortTP		3.514	-3.463	2.620	1.483	-1.250	-0.764	2.317
	PortFP			-5.230	-0.628	-1.828	-3.883	-3.492	-1.175
	PortFN				7.494	7.261	4.646	5.511	6.390
	PortTN					-2.692	-8.019	-7.796	-0.754
	HT-BM						-5.424	-4.621	1.592
	TF-BM							3.742	4.364
	VW-BM								3.775
	EW-BM								

Table 7: Fund performance of Deciles sorted by non-AI skills of fund managers

The non-AI skill level of a fund is defined as the fraction of stocks with positive *Trades* in the fund portfolio that are not selected by ANN. At the beginning of every quarter from 2013Q3 to 2022Q4, we sort mutual funds into deciles based on the non-AI skill level of a fund in the previous quarter. Decile 1 are funds with the lowest non-AI skill level while Decile 10 are funds with the highest non-AI skill level. A fund's total net asset value (TNA) at the beginning of each quarter is used in the calculation of value-weighted quarterly returns for each decile in the following quarter. This table reports the average of quarterly returns, average non-AI level, average proportion of TP, FP and TN stocks, average number of funds, the estimated coefficients, and the Newey-West (1987) t-statistics of the Carhart (1997) four-factor regression for the excess returns of the decile portfolios and of the long-short portfolio that buys Decile 10 and sells Decile 1. Panel A reports results of Gross return and Panel B presents results of Reported Net-of-expense Returns.

Panel A: Gross Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Return	0.0260	0.0260	0.0265	0.0292	0.0287	0.0264	0.0286	0.0288	0.0294	0.0291	0.0031
Average non-AI level	0.203	0.300	0.342	0.374	0.402	0.429	0.458	0.493	0.542	0.654	
Average TP proportion	0.078	0.074	0.067	0.056	0.049	0.042	0.038	0.035	0.030	0.022	
Average FP proportion	0.064	0.059	0.053	0.044	0.038	0.032	0.028	0.025	0.021	0.014	
Average TN proportion	0.655	0.567	0.538	0.525	0.512	0.497	0.476	0.447	0.407	0.310	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0057	-0.0050	-0.0043	-0.0011	-0.0021	-0.0039	-0.0012	0.0002	0.0005	0.0033	0.0090
	(-1.916)	(-1.933)	(-3.153)	(-0.938)	(-2.097)	(-1.655)	(-0.662)	(0.158)	(0.532)	(2.032)	(2.704)
Mkt-Rf	1.0105	0.9905	0.9870	0.9629	0.9964	0.9831	0.9674	0.9316	0.9289	0.8553	-0.1552
	(0.224)	(-0.269)	(-0.623)	(-1.291)	(-0.120)	(-0.442)	(-1.687)	(-2.545)	(-6.466)	(-6.778)	(-4.466)
SMB	0.0321	0.0309	0.0892	0.0758	0.0408	-0.0023	0.0876	0.0878	0.0942	0.1667	0.1346
	(0.705)	(0.691)	(2.735)	(3.250)	(1.342)	(-0.046)	(2.415)	(2.190)	(3.336)	(3.604)	(1.889)
HML	-0.0378	-0.0109	0.0613	0.0066	0.0083	0.0608	0.0305	0.0105	0.0067	0.0819	0.1198
	(-0.760)	(-0.259)	(2.217)	(0.156)	(0.216)	(0.995)	(0.660)	(0.257)	(0.246)	(2.190)	(1.704)
Momentum	0.1005	0.0869	0.0984	0.1146	0.0428	0.0339	0.0233	0.0092	0.0530	-0.0571	-0.1576
	(2.169)	(1.827)	(2.649)	(2.578)	(0.918)	(0.629)	(0.511)	(0.152)	(1.975)	(-1.553)	(-2.637)
Adjusted R2	0.9690	0.9767	0.9880	0.9859	0.9852	0.9727	0.9841	0.9828	0.9889	0.9805	0.4287

Table 7 (Cont.)

Panel B: Reported Net-of-expense Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Return	0.0240	0.0242	0.0249	0.0277	0.0272	0.0250	0.0271	0.0273	0.0278	0.0274	0.0035
Average non-AI level	0.203	0.300	0.342	0.374	0.402	0.429	0.458	0.493	0.542	0.654	
Average TP proportion	0.078	0.074	0.067	0.056	0.049	0.042	0.038	0.035	0.030	0.022	
Average FP proportion	0.064	0.059	0.053	0.044	0.038	0.032	0.028	0.025	0.021	0.014	
Average TN proportion	0.655	0.567	0.538	0.525	0.512	0.497	0.476	0.447	0.407	0.310	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0077	-0.0068	-0.0060	-0.0026	-0.0035	-0.0053	-0.0027	-0.0013	-0.0010	0.0017	0.0093
	(-2.609)	(-2.645)	(-4.314)	(-2.127)	(-3.541)	(-2.280)	(-1.528)	(-0.879)	(-1.109)	(1.050)	(2.826)
Mkt-Rf	1.0091	0.9896	0.9864	0.9618	0.9951	0.9820	0.9667	0.9310	0.9278	0.8544	-0.1546
	(0.192)	(-0.294)	(-0.641)	(-1.312)	(-0.164)	(-0.471)	(-1.718)	(-2.539)	(-6.428)	(-6.851)	(-4.427)
SMB	0.0334	0.0313	0.0894	0.0756	0.0418	-0.0013	0.0884	0.0881	0.0946	0.1668	0.1334
	(0.738)	(0.702)	(2.716)	(3.265)	(1.375)	(-0.025)	(2.442)	(2.203)	(3.383)	(3.654)	(1.890)
HML	-0.0366	-0.0093	0.0624	0.0072	0.0095	0.0620	0.0319	0.0119	0.0078	0.0831	0.1197
	(-0.729)	(-0.219)	(2.200)	(0.167)	(0.243)	(1.014)	(0.684)	(0.286)	(0.279)	(2.211)	(1.708)
Momentum	0.1019	0.0879	0.0995	0.1150	0.0447	0.0356	0.0249	0.0108	0.0536	-0.0557	-0.1575
	(2.148)	(1.826)	(2.580)	(2.513)	(0.939)	(0.654)	(0.539)	(0.177)	(1.962)	(-1.514)	(-2.650)
Adjusted R2	0.9686	0.9766	0.9877	0.9854	0.9849	0.9725	0.9840	0.9824	0.9889	0.9807	0.4298

Table 8: Performance of SVM portfolios and benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the SVM strategy and the benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). We report the performance of SVM portfolio, confusion matrix portfolios (PortTP, PortFP, PortFN and PortTN), and benchmark portfolios, including *Historical Trades Benchmark (HT-BM), Total Fund Benchmark (TF-BM), Market-Capitalization-Weight Benchmark (VW-BM), and Equal-Weight Benchmark (EW-BM)*.

		SVM Fund	PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	2.915	5.423	1.423	5.869	0.902	1.767	2.220	1.874	1.161
	Average quarterly return	0.042	0.055	0.035	0.053	0.020	0.030	0.033	0.030	0.026
	Volatility	0.127	0.130	0.190	0.084	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.165	-0.160	-0.204	-0.119	-0.170	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.318	0.413	0.177	0.608	0.216	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.245	0.335	0.164	0.429	0.106	0.184	0.200	0.183	0.128
10 bps	Total return	2.673	5.018	1.247	5.520	0.846	1.618	2.209	1.868	1.144
	Average quarterly return	0.040	0.054	0.033	0.051	0.019	0.028	0.033	0.030	0.025
	Volatility	0.127	0.130	0.190	0.084	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.166	-0.161	-0.206	-0.120	-0.171	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.305	0.401	0.167	0.592	0.208	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.234	0.322	0.154	0.413	0.101	0.173	0.199	0.182	0.127
50 bps	Total return	1.845	3.633	0.662	4.291	0.639	1.098	2.164	1.846	1.075
	Average quarterly return	0.034	0.047	0.026	0.046	0.016	0.023	0.033	0.030	0.025
	Volatility	0.126	0.129	0.188	0.083	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.171	-0.166	-0.212	-0.126	-0.173	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.255	0.351	0.127	0.530	0.172	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.188	0.272	0.113	0.352	0.082	0.132	0.196	0.181	0.122

Table 9: Performance of ANN_2 portfolios and benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN_2 strategy and the benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). We report the performance of ANN portfolio, confusion matrix portfolios (PortTN, PortFP, PortFN and PortTN), and benchmark portfolios, including *Historical Trades Benchmark (HT-BM), Total Fund Benchmark (TF-BM), Market-Capitalization-Weight Benchmark (VW-BM), and Equal-Weight Benchmark (EW-BM)*.

		ANN Fund	PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	3.855	4.863	0.663	5.610	0.869	1.767	2.220	1.874	1.161
	Average quarterly return	0.053	0.058	0.033	0.052	0.019	0.030	0.033	0.030	0.026
	Volatility	0.188	0.187	0.212	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.166	-0.162	-0.218	-0.123	-0.167	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.272	0.299	0.147	0.587	0.211	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.307	0.344	0.143	0.409	0.105	0.184	0.200	0.183	0.128
10 bps	Total return	3.555	4.508	0.541	5.304	0.814	1.618	2.209	1.868	1.144
	Average quarterly return	0.051	0.056	0.031	0.051	0.019	0.028	0.033	0.030	0.025
	Volatility	0.188	0.187	0.212	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.168	-0.164	-0.219	-0.124	-0.168	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.264	0.291	0.138	0.574	0.202	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.295	0.331	0.133	0.396	0.100	0.173	0.199	0.182	0.127
50 bps	Total return	2.524	3.287	0.135	4.211	0.607	1.098	2.164	1.846	1.075
	Average quarterly return	0.044	0.049	0.023	0.046	0.015	0.023	0.033	0.030	0.025
	Volatility	0.186	0.185	0.211	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.172	-0.169	-0.225	-0.128	-0.171	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.229	0.257	0.102	0.518	0.165	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.248	0.282	0.095	0.344	0.081	0.132	0.196	0.181	0.122

Table 10: Performance of ANN_32 portfolios and benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN_32 strategy and the benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). We report the performance of ANN portfolio, confusion matrix portfolios (PortTN, PortFP, PortFN and PortTN), and benchmark portfolios, including *Historical Trades Benchmark (HT-BM), Total Fund Benchmark (TF-BM), Market-Capitalization-Weight Benchmark (VW-BM), and Equal-Weight Benchmark (EW-BM)*.

		ANN Fund	PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	5.097	5.940	1.131	5.751	0.849	1.767	2.220	1.874	1.161
	Average quarterly return	0.059	0.059	0.041	0.052	0.019	0.030	0.033	0.030	0.026
	Volatility	0.182	0.145	0.240	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.205	-0.182	-0.236	-0.118	-0.167	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.314	0.397	0.163	0.595	0.208	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.278	0.317	0.166	0.430	0.104	0.184	0.200	0.183	0.128
10 bps	Total return	4.714	5.499	0.972	5.417	0.794	1.618	2.209	1.868	1.144
	Average quarterly return	0.057	0.058	0.039	0.051	0.018	0.028	0.033	0.030	0.025
	Volatility	0.182	0.145	0.239	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.206	-0.183	-0.237	-0.119	-0.168	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.305	0.386	0.154	0.581	0.199	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.269	0.306	0.156	0.415	0.099	0.173	0.199	0.182	0.127
50 bps	Total return	3.403	3.995	0.444	4.235	0.589	1.098	2.164	1.846	1.075
	Average quarterly return	0.050	0.051	0.031	0.046	0.015	0.023	0.033	0.030	0.025
	Volatility	0.180	0.143	0.238	0.084	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.210	-0.186	-0.243	-0.123	-0.170	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.270	0.341	0.122	0.521	0.162	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.231	0.262	0.119	0.357	0.079	0.132	0.196	0.181	0.122

Table 11: Performance of ANN_64_32_16_8 portfolios and benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN_64_32_16_8 strategy and the benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). We report the performance of ANN portfolio, confusion matrix portfolios (PortTP, PortFP, PortFN and PortTN), and benchmark portfolios, including *Historical Trades Benchmark (HT-BM), Total Fund Benchmark (TF-BM), Market-Capitalization-Weight Benchmark (VW-BM), and Equal-Weight Benchmark (EW-BM)*.

		ANN Fund	PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	3.289	1.448	1.238	5.587	0.849	1.767	2.220	1.874	1.161
	Average quarterly return	0.045	0.029	0.034	0.052	0.019	0.030	0.033	0.030	0.026
	Volatility	0.139	0.109	0.164	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.173	-0.203	-0.217	-0.121	-0.166	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.313	0.246	0.195	0.586	0.208	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.251	0.133	0.147	0.414	0.104	0.184	0.200	0.183	0.128
10 bps	Total return	3.033	1.300	1.076	5.285	0.793	1.618	2.209	1.868	1.144
	Average quarterly return	0.044	0.027	0.032	0.051	0.018	0.028	0.033	0.030	0.025
	Volatility	0.139	0.109	0.164	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.174	-0.204	-0.219	-0.122	-0.167	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.302	0.232	0.183	0.572	0.199	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.240	0.124	0.138	0.400	0.099	0.173	0.199	0.182	0.127
50 bps	Total return	2.149	0.791	0.535	4.206	0.587	1.098	2.164	1.846	1.075
	Average quarterly return	0.037	0.021	0.024	0.046	0.015	0.023	0.033	0.030	0.025
	Volatility	0.138	0.109	0.163	0.085	0.083	0.088	0.083	0.081	0.113
	5% VaR	-0.179	-0.208	-0.225	-0.126	-0.169	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.257	0.174	0.137	0.517	0.162	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.198	0.091	0.099	0.349	0.079	0.132	0.196	0.181	0.122

Table 12: Fund performance of Deciles sorted by non-AI skills of fund managers: different machine learning models

The non-AI skill level of a fund is defined as the fraction of stocks with positive *Trades* in the fund portfolio that are not selected by a machine learning model. This table present the result for four models: SVM, ANN_2, ANN_32, and ANN_64_32_16_8. At the beginning of every quarter from 2013Q3 through 2022Q4, we sort mutual funds into deciles based on the non-AI skill level of a fund with a one-quarter lag. Decile 1 are funds with the lowest non-AI skill level while Decile 10 are funds with the highest non-AI skill level. A fund's total net asset value (TNA) at the beginning of each quarter is used in the calculation of value-weighted quarterly returns for each decile in the following quarter. This table reports the average of quarterly returns, average non-AI level, average proportion of TP, FP and TN stocks, average number of funds, the estimated coefficients, and the Newey-West (1987) t-statistics of the Carhart (1997) four-factor regression for the excess returns of the decile portfolios and of the long-short portfolio that buys Decile 10 and sells Decile 1.

Panel A: SVM

					Gross Return						
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0242	0.0250	0.0250	0.0268	0.0282	0.0263	0.0315	0.0290	0.0273	0.0310	0.0067
Average non-AI level	0.184	0.279	0.320	0.349	0.375	0.401	0.428	0.462	0.509	0.620	
Average TP prop	0.103	0.096	0.089	0.085	0.080	0.073	0.070	0.064	0.058	0.045	
Average FP prop	0.082	0.076	0.072	0.067	0.062	0.057	0.052	0.048	0.041	0.032	
Average TN prop	0.631	0.549	0.520	0.499	0.484	0.469	0.449	0.427	0.391	0.303	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0081	-0.0071	-0.0069	-0.0051	-0.0028	-0.0042	0.0026	0.0006	-0.0007	0.0062	0.0143
	(-2.859)	(-2.735)	(-2.524)	(-4.800)	(-1.801)	(-2.664)	(1.707)	(0.531)	(-0.646)	(3.819)	(4.376)
Mkt-Rf	1.0355	1.0239	1.0241	1.0137	1.0093	0.9910	0.9370	0.9223	0.9017	0.8224	-0.2130
	(0.905)	(0.811)	(0.607)	(0.666)	(0.300)	(-0.388)	(-2.241)	(-3.678)	(-7.299)	(-6.720)	(-4.887)
SMB	0.0462	0.0344	0.0363	0.1016	0.0572	0.0401	0.0834	0.1043	0.0516	0.1670	0.1208
	(1.206)	(0.823)	(0.937)	(4.090)	(1.350)	(1.266)	(1.959)	(2.863)	(1.486)	(2.921)	(1.698)
HML	-0.0432	0.0115	0.0542	0.0681	0.0163	0.0497	0.0415	0.0534	0.0038	0.0240	0.0672
	(-0.706)	(0.184)	(0.747)	(1.718)	(0.413)	(1.105)	(0.932)	(1.312)	(0.163)	(0.566)	(0.786)
Momentum	0.0893	0.1108	0.0883	0.1403	0.0313	0.0329	0.0370	0.0238	0.0367	-0.0732	-0.1625
	(1.616)	(2.335)	(1.875)	(4.684)	(0.564)	(0.793)	(0.673)	(0.505)	(1.483)	(-1.634)	(-2.554)
Adjusted R2	0.9669	0.9736	0.9666	0.9877	0.9808	0.9858	0.9763	0.9852	0.9880	0.9685	0.3339
					Reported Net-of-expense Return						
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0223	0.0231	0.0234	0.0253	0.0268	0.0248	0.0300	0.0275	0.0257	0.0293	0.0071
Average non-AI level	0.184	0.279	0.320	0.349	0.375	0.401	0.428	0.462	0.509	0.620	
Average TP prop	0.103	0.096	0.089	0.085	0.080	0.073	0.070	0.064	0.058	0.045	
Average FP prop	0.082	0.076	0.072	0.067	0.062	0.057	0.052	0.048	0.041	0.032	
Average TN prop	0.631	0.549	0.520	0.499	0.484	0.469	0.449	0.427	0.391	0.303	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0100	-0.0090	-0.0084	-0.0066	-0.0043	-0.0057	0.0011	-0.0009	-0.0022	0.0046	0.0146
	(-3.613)	(-3.484)	(-3.153)	(-6.443)	(-2.702)	(-3.701)	(0.726)	(-0.895)	(-2.043)	(2.851)	(4.533)
Mkt-Rf	1.0341	1.0233	1.0232	1.0132	1.0080	0.9903	0.9358	0.9213	0.9010	0.8211	-0.2130
	(0.869)	(0.790)	(0.588)	(0.643)	(0.258)	(-0.421)	(-2.266)	(-3.709)	(-7.307)	(-6.728)	(-4.888)
SMB	0.0477	0.0346	0.0361	0.1013	0.0581	0.0411	0.0839	0.1051	0.0517	0.1673	0.1196
	(1.247)	(0.828)	(0.937)	(4.134)	(1.369)	(1.317)	(1.975)	(2.901)	(1.507)	(2.961)	(1.702)
HML	-0.0419	0.0130	0.0550	0.0685	0.0185	0.0507	0.0425	0.0544	0.0054	0.0249	0.0668
	(-0.684)	(0.209)	(0.760)	(1.719)	(0.465)	(1.122)	(0.935)	(1.318)	(0.226)	(0.580)	(0.784)
Momentum	0.0909	0.1119	0.0890	0.1408	0.0340	0.0344	0.0386	0.0249	0.0378	-0.0723	-0.1631
	(1.616)	(2.335)	(1.863)	(4.541)	(0.603)	(0.820)	(0.688)	(0.523)	(1.533)	(-1.615)	(-2.592)
Adjusted R2	0.9666	0.9736	0.9665	0.9876	0.9806	0.9858	0.9759	0.9851	0.9880	0.9686	0.3349

Table 12 (Cont.)

Panel B: ANN_2

Gross Return

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0243	0.0248	0.0267	0.0285	0.0261	0.0278	0.0292	0.0302	0.0296	0.0291	0.0048
Average non-AI level	0.196	0.292	0.335	0.367	0.395	0.422	0.451	0.487	0.537	0.650	
Average TP prop	0.090	0.085	0.076	0.064	0.055	0.048	0.043	0.039	0.033	0.025	
Average FP prop	0.071	0.066	0.058	0.051	0.042	0.037	0.031	0.028	0.023	0.016	
Average TN prop	0.643	0.557	0.531	0.518	0.508	0.493	0.475	0.446	0.408	0.309	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0088	-0.0074	-0.0052	-0.0028	-0.0038	-0.0028	-0.0004	0.0006	0.0014	0.0046	0.0134
	(-2.741)	(-3.303)	(-3.689)	(-2.277)	(-3.061)	(-2.313)	(-0.218)	(0.401)	(1.137)	(2.548)	(3.451)
Mkt-Rf	1.0478	1.0204	1.0098	0.9924	0.9576	0.9962	0.9525	0.9675	0.9032	0.8233	-0.2245
	(0.830)	(0.489)	(0.477)	(-0.309)	(-1.376)	(-0.172)	(-2.391)	(-1.459)	(-4.368)	(-7.595)	(-3.377)
SMB	0.0501	0.0833	0.1276	0.0737	0.0582	0.0135	0.0632	0.1009	0.0888	0.1474	0.0973
	(1.130)	(2.661)	(3.605)	(2.250)	(1.739)	(0.335)	(1.917)	(2.575)	(2.832)	(2.734)	(1.161)
HML	-0.0217	-0.0050	0.0645	0.0338	0.0007	0.0620	0.0674	0.0109	0.0407	0.0389	0.0605
	(-0.456)	(-0.143)	(2.537)	(0.809)	(0.014)	(1.801)	(1.181)	(0.234)	(1.481)	(1.172)	(0.951)
Momentum	0.1465	0.1397	0.1430	0.1336	0.0777	0.0226	0.0504	0.0032	0.0530	-0.1161	-0.2626
	(3.352)	(3.250)	(3.584)	(2.608)	(1.578)	(0.541)	(1.162)	(0.058)	(1.550)	(-2.960)	(-3.856)
Adjusted R2	0.9664	0.9770	0.9873	0.9841	0.9785	0.9864	0.9793	0.9827	0.9871	0.9773	0.4421

Reported Net-of-expense Return

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0223	0.0230	0.0251	0.0269	0.0246	0.0264	0.0277	0.0287	0.0280	0.0275	0.0052
Average non-AI level	0.196	0.292	0.335	0.367	0.395	0.422	0.451	0.487	0.537	0.650	
Average TP prop	0.090	0.085	0.076	0.064	0.055	0.048	0.043	0.039	0.033	0.025	
Average FP prop	0.071	0.066	0.058	0.051	0.042	0.037	0.031	0.028	0.023	0.016	
Average TN prop	0.643	0.557	0.531	0.518	0.508	0.493	0.475	0.446	0.408	0.309	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0108	-0.0092	-0.0068	-0.0043	-0.0053	-0.0042	-0.0018	-0.0009	-0.0001	0.0030	0.0138
	(-3.391)	(-4.142)	(-4.863)	(-3.464)	(-4.342)	(-3.556)	(-1.133)	(-0.610)	(-0.056)	(1.672)	(3.558)
Mkt-Rf	1.0465	1.0195	1.0096	0.9912	0.9565	0.9953	0.9517	0.9664	0.9021	0.8226	-0.2240
	(0.805)	(0.465)	(0.459)	(-0.351)	(-1.403)	(-0.211)	(-2.418)	(-1.491)	(-4.422)	(-7.594)	(-3.368)
SMB	0.0513	0.0834	0.1285	0.0733	0.0594	0.0151	0.0632	0.1012	0.0890	0.1478	0.0965
	(1.170)	(2.659)	(3.648)	(2.240)	(1.790)	(0.376)	(1.922)	(2.586)	(2.882)	(2.754)	(1.158)
HML	-0.0204	-0.0034	0.0660	0.0343	0.0020	0.0632	0.0685	0.0124	0.0417	0.0402	0.0606
	(-0.429)	(-0.097)	(2.525)	(0.802)	(0.038)	(1.812)	(1.192)	(0.262)	(1.496)	(1.203)	(0.957)
Momentum	0.1479	0.1405	0.1452	0.1336	0.0796	0.0249	0.0519	0.0041	0.0535	-0.1140	-0.2619
	(3.309)	(3.212)	(3.548)	(2.553)	(1.582)	(0.590)	(1.172)	(0.073)	(1.547)	(-2.884)	(-3.843)
Adjusted R2	0.9662	0.9768	0.9870	0.9837	0.9781	0.9863	0.9791	0.9824	0.9871	0.9774	0.4428

Table 12 (Cont.)

Panel C: ANN_32

Gross Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0264	0.0266	0.0288	0.0288	0.0278	0.0287	0.0284	0.0289	0.0290	0.0269	0.0005
Average non-AI level	0.179	0.273	0.315	0.346	0.373	0.399	0.428	0.463	0.511	0.624	
Average TP prop	0.114	0.108	0.096	0.086	0.079	0.071	0.067	0.062	0.055	0.042	
Average FP prop	0.099	0.090	0.082	0.074	0.067	0.060	0.055	0.050	0.043	0.032	
Average TN prop	0.608	0.529	0.507	0.494	0.481	0.469	0.450	0.425	0.392	0.303	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0052	-0.0053	-0.0019	-0.0016	-0.0017	-0.0019	-0.0027	0.0002	0.0013	0.0012	0.0064
	(-2.480)	(-2.962)	(-1.789)	(-1.494)	(-1.645)	(-1.719)	(-1.161)	(0.178)	(0.743)	(0.744)	(2.747)
Mkt-Rf	1.0012	1.0202	0.9800	0.9667	0.9582	0.9865	1.0070	0.9321	0.9062	0.8454	-0.1559
	(0.025)	(0.484)	(-0.696)	(-1.327)	(-1.106)	(-0.618)	(0.190)	(-2.683)	(-3.534)	(-6.354)	(-2.471)
SMB	0.0872	0.0288	0.0819	0.1285	0.0403	0.0646	0.0875	0.0516	0.0603	0.1254	0.0382
	(2.006)	(0.697)	(3.847)	(4.126)	(1.474)	(2.085)	(1.871)	(1.470)	(1.583)	(2.426)	(0.513)
HML	-0.0841	-0.0331	0.0216	0.0218	0.0406	0.0058	0.0679	0.0240	0.0303	0.0532	0.1373
	(-2.032)	(-1.142)	(0.623)	(0.612)	(0.981)	(0.195)	(1.287)	(0.751)	(0.926)	(1.326)	(2.944)
Momentum	0.1191	0.1024	0.1002	0.1197	0.0208	0.0573	0.0551	0.0171	-0.0107	-0.0361	-0.1553
	(2.344)	(2.413)	(2.340)	(2.688)	(0.395)	(1.379)	(1.129)	(0.380)	(-0.227)	(-0.991)	(-2.382)
Adjusted R2	0.9713	0.9751	0.9858	0.9848	0.9831	0.9879	0.9796	0.9844	0.9850	0.9714	0.4149
Reported Net-of-expense Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0244	0.0248	0.0272	0.0273	0.0263	0.0272	0.0269	0.0274	0.0275	0.0253	0.0010
Average non-AI level	0.179	0.273	0.315	0.346	0.373	0.399	0.428	0.463	0.511	0.624	
Average TP prop	0.114	0.108	0.096	0.086	0.079	0.071	0.067	0.062	0.055	0.042	
Average FP prop	0.099	0.090	0.082	0.074	0.067	0.060	0.055	0.050	0.043	0.032	
Average TN prop	0.608	0.529	0.507	0.494	0.481	0.469	0.450	0.425	0.392	0.303	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0072	-0.0071	-0.0035	-0.0031	-0.0032	-0.0033	-0.0041	-0.0013	-0.0002	-0.0004	0.0068
	(-3.457)	(-4.006)	(-3.298)	(-2.833)	(-3.078)	(-3.007)	(-1.807)	(-1.095)	(-0.095)	(-0.240)	(2.913)
Mkt-Rf	0.9999	1.0193	0.9792	0.9661	0.9575	0.9857	1.0054	0.9311	0.9051	0.8445	-0.1554
	(-0.001)	(0.462)	(-0.722)	(-1.336)	(-1.118)	(-0.647)	(0.147)	(-2.712)	(-3.568)	(-6.407)	(-2.465)
SMB	0.0882	0.0292	0.0812	0.1291	0.0406	0.0653	0.0892	0.0521	0.0612	0.1254	0.0373
	(2.037)	(0.718)	(3.877)	(4.105)	(1.491)	(2.104)	(1.909)	(1.495)	(1.612)	(2.457)	(0.507)
HML	-0.0828	-0.0315	0.0226	0.0237	0.0423	0.0070	0.0689	0.0253	0.0313	0.0542	0.1370
	(-1.969)	(-1.066)	(0.632)	(0.651)	(1.012)	(0.228)	(1.290)	(0.777)	(0.944)	(1.357)	(2.950)
Momentum	0.1206	0.1038	0.1007	0.1214	0.0228	0.0587	0.0572	0.0183	-0.0099	-0.0351	-0.1557
	(2.320)	(2.416)	(2.275)	(2.655)	(0.427)	(1.383)	(1.151)	(0.403)	(-0.208)	(-0.971)	(-2.405)
Adjusted R2	0.9709	0.9750	0.9854	0.9844	0.9829	0.9876	0.9793	0.9844	0.9848	0.9718	0.4166

Table 12 (Cont.)

Panel D: ANN_64_32_16_8

	Gross Return										
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0247	0.0258	0.0263	0.0276	0.0273	0.0267	0.0309	0.0315	0.0270	0.0297	0.0050
Average non-AI level	0.191	0.287	0.330	0.363	0.391	0.419	0.448	0.484	0.533	0.648	
Average TP prop	0.094	0.087	0.077	0.068	0.060	0.054	0.048	0.043	0.038	0.027	
Average FP prop	0.076	0.069	0.062	0.055	0.047	0.043	0.037	0.032	0.027	0.019	
Average TN prop	0.639	0.557	0.530	0.514	0.502	0.485	0.467	0.441	0.402	0.306	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0074	-0.0056	-0.0046	-0.0018	-0.0042	-0.0041	0.0021	0.0029	-0.0016	0.0036	0.0110
	(-2.668)	(-2.760)	(-2.863)	(-1.472)	(-3.674)	(-2.002)	(1.352)	(2.187)	(-1.040)	(2.710)	(3.987)
Mkt-Rf	1.0180	0.9993	0.9936	0.9315	1.0106	0.9986	0.9365	0.9252	0.9254	0.8646	-0.1535
	(0.338)	(-0.024)	(-0.261)	(-1.752)	(0.538)	(-0.054)	(-3.063)	(-2.684)	(-2.609)	(-5.688)	(-2.970)
SMB	0.0629	0.0660	0.1042	0.0541	0.0957	0.0373	0.0770	0.1188	0.0266	0.1426	0.0797
	(1.155)	(1.478)	(3.209)	(1.437)	(2.882)	(0.872)	(1.878)	(2.539)	(0.713)	(3.167)	(1.023)
HML	-0.1077	-0.0532	0.0049	0.0093	0.0214	0.0459	0.0787	0.0444	0.0232	0.0695	0.1773
	(-2.182)	(-1.101)	(0.133)	(0.237)	(0.669)	(1.077)	(1.766)	(1.089)	(1.090)	(1.873)	(3.241)
Momentum	0.1138	0.1080	0.0780	0.1033	0.0900	0.0432	0.0184	0.0345	0.0189	-0.0612	-0.1750
	(2.118)	(1.996)	(1.831)	(2.862)	(1.826)	(0.901)	(0.391)	(0.641)	(0.623)	(-1.661)	(-2.936)
Adjusted R2	0.9660	0.9737	0.9843	0.9801	0.9867	0.9815	0.9844	0.9802	0.9902	0.9796	0.5126
	Reported Net-of-expense Return										
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0226	0.0239	0.0247	0.0261	0.0258	0.0252	0.0294	0.0299	0.0255	0.0281	0.0054
Average non-AI level	0.191	0.287	0.330	0.363	0.391	0.419	0.448	0.484	0.533	0.648	
Average TP prop	0.094	0.087	0.077	0.068	0.060	0.054	0.048	0.043	0.038	0.027	
Average FP prop	0.076	0.069	0.062	0.055	0.047	0.043	0.037	0.032	0.027	0.019	
Average TN prop	0.639	0.557	0.530	0.514	0.502	0.485	0.467	0.441	0.402	0.306	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0094	-0.0074	-0.0062	-0.0033	-0.0056	-0.0056	0.0007	0.0014	-0.0031	0.0020	0.0114
	(-3.419)	(-3.706)	(-3.951)	(-2.670)	(-4.869)	(-2.751)	(0.420)	(1.025)	(-2.079)	(1.557)	(4.161)
Mkt-Rf	1.0170	0.9986	0.9927	0.9311	1.0095	0.9970	0.9356	0.9242	0.9247	0.8635	-0.1535
	(0.317)	(-0.047)	(-0.298)	(-1.746)	(0.480)	(-0.112)	(-3.089)	(-2.688)	(-2.629)	(-5.685)	(-2.964)
SMB	0.0642	0.0660	0.1042	0.0549	0.0956	0.0387	0.0772	0.1198	0.0268	0.1426	0.0785
	(1.183)	(1.477)	(3.180)	(1.463)	(2.871)	(0.907)	(1.889)	(2.557)	(0.726)	(3.197)	(1.013)
HML	-0.1061	-0.0517	0.0061	0.0101	0.0228	0.0473	0.0795	0.0459	0.0243	0.0706	0.1767
	(-2.137)	(-1.062)	(0.163)	(0.252)	(0.705)	(1.106)	(1.758)	(1.097)	(1.137)	(1.869)	(3.246)
Momentum	0.1156	0.1089	0.0787	0.1046	0.0913	0.0453	0.0189	0.0365	0.0195	-0.0600	-0.1756
	(2.116)	(1.978)	(1.800)	(2.825)	(1.825)	(0.934)	(0.395)	(0.667)	(0.645)	(-1.600)	(-2.964)
Adjusted R2	0.9657	0.9734	0.9840	0.9798	0.9865	0.9814	0.9842	0.9798	0.9903	0.9796	0.5134

Table 13: Performance of ANN_TopNonAI portfolios and benchmarks

This table shows the performance measures of portfolios held from 2013Q1–2022Q4 and rebalanced quarterly using the ANN_TopNonAI strategy and the benchmarks. Total return is the compound returns of a portfolio over time. Average quarterly return is the simple average of the quarterly return of a portfolio. Volatility is the standard deviation of the quarterly return of a portfolio. VaR is the Value-at-Risk of the quarterly return of a portfolio at the 5 percent significance level. Sharpe_Volatility of a portfolio is its average return in excess of the average three-month T-Bill rate over the portfolio's volatility from 2013Q1–2022Q4. Sharpe_VaR of a portfolio is its average return in excess of the average three-month T-Bill rate over the absolute value of the portfolio's 5 percent VaR from 2013Q1–2022Q4. The results are reported for three levels of trading costs: 0 bp (frictionless), 10bps (moderate transaction cost), and 50bps (high transaction cost). We report the performance of ANN portfolio, confusion matrix portfolios (PortTP, PortFP, PortFN and PortTN), and benchmark portfolios, including *Historical Trades Benchmark (HT-BM), Total Fund Benchmark (TF-BM), Market-Capitalization-Weight Benchmark (VW-BM), and Equal-Weight Benchmark (EW-BM)*.

		ANN Fund	PortTP	PortFP	PortFN	PortTN	HT-BM	TF-BM	VW-BM	EW-BM
0 bp	Total return	0.881	6.382	0.556	5.948	0.914	1.767	2.220	1.874	1.161
	Average quarterly return	0.023	0.060	0.029	0.053	0.020	0.030	0.033	0.030	0.026
	Volatility	0.116	0.139	0.214	0.085	0.084	0.088	0.083	0.081	0.113
	5% VaR	-0.238	-0.193	-0.246	-0.122	-0.166	-0.152	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.186	0.419	0.126	0.607	0.217	0.316	0.379	0.350	0.212
	Sharpe_VaR	0.090	0.302	0.109	0.423	0.110	0.184	0.200	0.183	0.128
10 bps	Total return	0.773	5.904	0.448	5.626	0.863	1.618	2.209	1.868	1.144
	Average quarterly return	0.022	0.058	0.027	0.052	0.019	0.028	0.033	0.030	0.025
	Volatility	0.116	0.139	0.213	0.084	0.084	0.088	0.083	0.081	0.113
	5% VaR	-0.239	-0.194	-0.248	-0.122	-0.166	-0.153	-0.157	-0.155	-0.187
	Sharpe_Volatility	0.173	0.407	0.117	0.594	0.209	0.301	0.378	0.350	0.210
	Sharpe_VaR	0.084	0.291	0.101	0.409	0.105	0.173	0.199	0.182	0.127
50 bps	Total return	0.400	4.274	0.085	4.476	0.671	1.098	2.164	1.846	1.075
	Average quarterly return	0.016	0.051	0.019	0.047	0.016	0.023	0.033	0.030	0.025
	Volatility	0.115	0.138	0.212	0.084	0.084	0.088	0.083	0.081	0.113
	5% VaR	-0.245	-0.200	-0.254	-0.126	-0.169	-0.158	-0.158	-0.155	-0.188
	Sharpe_Volatility	0.122	0.359	0.084	0.538	0.176	0.237	0.374	0.347	0.203
	Sharpe_VaR	0.057	0.247	0.070	0.358	0.088	0.132	0.196	0.181	0.122

Table 14: Fund performance of Deciles sorted by non-AI skills of fund managers: ANN_TopNonAI

The non-AI skill level of a fund is defined as the fraction of stocks with positive *Trades* in the fund portfolio that are not selected by ANN_TopNonAI. At the beginning of every quarter from 2013Q3 to 2022Q4, we sort mutual funds into deciles based on the non-AI skill level of a fund in the previous quarter. Decile 1 are funds with the lowest non-AI skill level while Decile 10 are funds with the highest non-AI skill level. A fund's total net asset value (TNA) at the beginning of each quarter is used in the calculation of value-weighted quarterly returns for each decile in the following quarter. This table reports the average of quarterly returns, average non-AI level, average proportion of TP, FP and TN stocks, average number of funds, the estimated coefficients, and the Newey-West (1987) t-statistics of the Carhart (1997) four-factor regression for the excess returns of the decile portfolios and of the long-short portfolio that buys Decile 10 and sells Decile 1. Panel A reports results of Gross return and Panel B presents results of Reported Net-of-expense Returns.

Panel A: Gross Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0252	0.0261	0.0297	0.0291	0.0271	0.0267	0.0283	0.0282	0.0303	0.0257	0.0005
Average non-AI level	0.144	0.233	0.277	0.311	0.342	0.372	0.404	0.443	0.496	0.612	
Average TP proportion	0.117	0.110	0.099	0.088	0.079	0.071	0.064	0.057	0.051	0.038	
Average FP proportion	0.128	0.115	0.101	0.088	0.077	0.068	0.060	0.052	0.047	0.035	
Average TN proportion	0.611	0.541	0.522	0.512	0.503	0.490	0.472	0.448	0.407	0.315	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0060	-0.0057	-0.0010	-0.0014	-0.0044	-0.0040	-0.0014	-0.0024	0.0051	-0.0017	0.0043
	(-1.619)	(-2.741)	(-0.433)	(-1.128)	(-3.083)	(-2.971)	(-1.005)	(-1.099)	(4.599)	(-0.627)	(0.812)
Mkt-Rf	0.9923	1.0135	0.9755	0.9777	1.0090	0.9801	0.9638	0.9940	0.8349	0.9129	-0.0794
	(-0.125)	(0.453)	(-0.565)	(-0.772)	(0.292)	(-0.696)	(-1.583)	(-0.195)	(-11.896)	(-2.334)	(-0.972)
SMB	0.0910	0.1202	0.1200	0.0926	0.0860	0.1024	0.0851	0.0479	0.1019	0.0683	-0.0227
	(1.615)	(2.912)	(3.405)	(2.855)	(2.398)	(3.345)	(2.351)	(1.059)	(3.136)	(1.793)	(-0.347)
HML	-0.0547	0.0617	0.0552	0.0704	0.0917	0.0473	0.0392	0.0589	-0.0117	0.0006	0.0552
	(-0.912)	(1.727)	(1.088)	(1.835)	(2.339)	(1.136)	(0.836)	(1.274)	(-0.554)	(0.010)	(0.548)
Momentum	0.1091	0.1329	0.1229	0.0915	0.1013	0.1047	0.0309	0.0346	-0.0729	-0.0778	-0.1869
	(1.422)	(2.716)	(2.364)	(1.832)	(2.559)	(1.807)	(0.636)	(0.613)	(-2.227)	(-1.646)	(-1.736)
Adjusted R2	0.9518	0.9751	0.9720	0.9819	0.9816	0.9817	0.9817	0.9832	0.9881	0.9714	0.0998

Table 14 (Cont.)

Panel B: Reported Net-of-expense Return											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10-D1
Average Returns	0.0232	0.0243	0.0281	0.0276	0.0257	0.0252	0.0268	0.0267	0.0289	0.0240	0.0009
Average non-AI level	0.144	0.233	0.277	0.311	0.342	0.372	0.404	0.443	0.496	0.612	
Average TP proportion	0.117	0.110	0.099	0.088	0.079	0.071	0.064	0.057	0.051	0.038	
Average FP proportion	0.128	0.115	0.101	0.088	0.077	0.068	0.060	0.052	0.047	0.035	
Average TN proportion	0.611	0.541	0.522	0.512	0.503	0.490	0.472	0.448	0.407	0.315	
Average No. of Funds	250	250	250	250	250	250	250	250	250	254	
Alpha	-0.0080	-0.0075	-0.0026	-0.0029	-0.0058	-0.0055	-0.0029	-0.0039	0.0037	-0.0033	0.0047
	(-2.166)	(-3.587)	(-1.102)	(-2.358)	(-4.148)	(-4.009)	(-2.072)	(-1.776)	(3.298)	(-1.214)	(0.883)
Mkt-Rf	0.9909	1.0131	0.9745	0.9768	1.0081	0.9792	0.9633	0.9928	0.8337	0.9119	-0.0790
	(-0.148)	(0.434)	(-0.580)	(-0.801)	(0.263)	(-0.724)	(-1.591)	(-0.237)	(-11.932)	(-2.355)	(-0.961)
SMB	0.0929	0.1196	0.1211	0.0928	0.0861	0.1032	0.0854	0.0495	0.1024	0.0683	-0.0246
	(1.665)	(2.886)	(3.456)	(2.882)	(2.397)	(3.363)	(2.366)	(1.097)	(3.166)	(1.790)	(-0.378)
HML	-0.0528	0.0624	0.0561	0.0714	0.0928	0.0485	0.0400	0.0609	-0.0103	0.0022	0.0551
	(-0.868)	(1.700)	(1.084)	(1.828)	(2.321)	(1.139)	(0.843)	(1.310)	(-0.477)	(0.037)	(0.544)
Momentum	0.1106	0.1334	0.1245	0.0925	0.1030	0.1059	0.0322	0.0359	-0.0718	-0.0765	-0.1871
	(1.427)	(2.654)	(2.338)	(1.820)	(2.547)	(1.800)	(0.652)	(0.634)	(-2.194)	(-1.615)	(-1.736)
Adjusted R2	0.9511	0.9746	0.9714	0.9817	0.9814	0.9813	0.9815	0.9831	0.9880	0.9715	0.0982