The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1985-9899.htm

Stock price indices prediction combining deep learning algorithms and selected technical indicators based on correlation

Abdelhadi Ifleh and Mounime El Kabbouri
Department of Finance, Audit and Organizational Governance Research Laboratory, National School of Commerce and Management, Hassan First University of Settat, Settat, Morocco

Received 18 February 2023
Revised 16 April 2023
17 August 2023
Accepted 28 September 2023

Abstract

Purpose – The prediction of stock market (SM) indices is a fascinating task. An in-depth analysis in this field can provide valuable information to investors, traders and policy makers in attractive SMs. This article aims to apply a correlation feature selection model to identify important technical indicators (TIs), which are combined with multiple deep learning (DL) algorithms for forecasting SM indices.

Design/methodology/approach – The methodology involves using a correlation feature selection model to select the most relevant features. These features are then used to predict the fluctuations of six markets using various DL algorithms, and the results are compared with predictions made using all features by using a range of performance measures.

Findings – The experimental results show that the combination of TIs selected through correlation and Artificial Neural Network (ANN) provides good results in the MADEX market. The combination of selected indicators and Convolutional Neural Network (CNN) in the NASDAQ 100 market outperforms all other combinations of variables and models. In other markets, the combination of all variables with ANN provides the best results.

Originality/value – This article makes several significant contributions, including the use of a correlation feature selection model to select pertinent variables, comparison between multiple DL algorithms (ANN, CNN and Long-Short-Term Memory (LSTM)), combining selected variables with algorithms to improve predictions, evaluation of the suggested model on six datasets (MASI, MADEX, FTSE 100, SP500, NASDAQ 100 and EGX 30) and application of various performance measures (Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error(RMSE), Mean Squared Logarithmic Error (MSLE) and Root Mean Squared Logarithmic Error (RMSLE)).

Keywords Stock market indices, DL, EMH, Performance measurement, Feature selection

Paper type Research paper

1. Introduction

SM prediction is a long-standing and difficult task due to the inherent complexities of financial time series, such as high volatility, non-stationarity and non-linearity (Long, Chen, He, Wu, & Ren, 2019). The Efficient Market Hypothesis posits that it is impossible to predict stock price movements and prices behave randomly (Fama, 1965). In contrast, Technical Analysis (TA) claims that prices incorporate all available information and trend detection makes price prediction easier (Patel, 2014).

© Abdelhadi Ifleh and Mounime El Kabbouri. Published in *Arab Gulf Journal of Scientific Research*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

Arab Gulf Journal of Scientific Research
Vol. 42 No. 4, 2024
pp. 1237-1256
Emerald Publishing Limited
e-ISSN: 2536-0051
p-ISSN: 1985-9899
DOI 10.1108/AGJSR-02-2023-0070

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Investment decisions in financial markets can be made through either fundamental analysis or TA. Fundamental analysis involves evaluating the actual price against the intrinsic value and deciding to buy or sell based on this comparison. TA, on the other hand, relies on historical data and employs TIs to help traders determine when to buy and sell assets (Naik & Mohan, 2019; Ratto, Merello, Ma, Oneto, & Cambria, 2019).

In recent times, various studies have combined Artificial Intelligence (AI) algorithms with TIs for more accurate financial market predictions. The most commonly used models include Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) (Sezer, Gudelek, & Ozbayoglu, 2019). This study aims to predict different SM indices by utilizing ANN, CNN and LSTM combined with TIs. Correlation is employed as the feature selection method for selecting relevant TIs. Related works are discussed in the following section, followed by a description of our methodology and accuracy metrics in Section 3. Our findings are analyzed in Section 4, and the conclusion, along with a discussion of future work, is provided in Section 5.

2. Literature review

The growth of computer technology has made it easier to use and develop TA in the SM. With the help of these tools, investors are now able to create powerful decision support systems that can increase their profits and minimize losses.

Chopra, Yadav, and Chopra (2019) employed an ANN model to predict the stock prices of nine companies with diverse market capitalizations and the CNX NIFTY50 index on the Indian Stock Exchange. Their research highlights the model's effectiveness in predicting stock prices, especially for the most volatile prices before and after the demonetization process.

Dash and Dash (2016) in their study, presented a hybrid stock trading framework that combines TA with machine learning techniques. They proposed a decision support system called Computational Efficient Functional Link Artificial Neural Network (CEFLANN) to help investors make more informed decisions with less risk. The system uses an Extreme Learning Machine (ELM) to generate investment decisions and was compared to other models such as Support Vector Machine (SVM), Naive Bayesian, K Nearest Neighbor (KNN) and Decision Tree (DT). The results showed that CEFLANN is more profitable compared to these other models.

Agrawal, Khan, and Shukla (2019) developed a model that forecasts SM price movements based on TIs. They utilized the "Optimal Long Short-Term Memory (O_LSTM)" model, which was able to predict both short-term and long-term trends. The results of this model showed its superior performance compared to other models like ELSTM, SVM and linear regression (LR).

Selvamuthu, Kumar, and Mishra (2019) compared the predictive abilities of three neural network learning algorithms: Levenberg-Marquardt, Scaled Conjugate Gradient and Bayesian Regularization. The results of this study revealed the predictive accuracy of these three algorithms with a score of 99.9%.

Qiu and Song (2016) used the "Artificial Neural Network (ANN)" model combined with the "Genetic Algorithm (GA)" model. Also, they introduced two types of inputs in the form of the TA indicators. The results showed the effectiveness of this hybrid model in the daily forecasting of the Nikkie 225 index compared to other studies that used other models.

Qiu and Song (2016) employed an ANN model in conjunction with a "Genetic Algorithm (GA)" model and introduced two types of inputs in the form of TA indicators. Their results showed that this hybrid model was effective in daily forecasting of the Nikkei 225 index compared to other studies that used alternative models.

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Efat, Bashar, Imtiaz Ud-Din, and Bhuiyan (2018) utilized a new model called "Trend Estimation with Linear Regression" to predict market trends. When comparing their findings with those from the ARIMA and PROPHET models, their results were more favorable.

Sezer, Ozbayoglu, and Dogdu (2017) developed a decision support system for investors to determine entry points, using neural networks applied to the Relative Strength Index (RSI) and moving average.

Sahoo and Mohanty (2020) combined an ANN model with "The Gray Wolf Optimization" to forecast prices on the Bombay Stock Exchange. They found that this combination works better than an ANN model alone.

Sang and Di Pierro (2019) examined the application of machine learning techniques in stock trading and proposed the use of a LSTM neural network to enhance the accuracy of predictions made with TA. The authors tested their proposed method using historical stock data and found that it outperforms traditional TA methods. Their research concludes that incorporating LSTM with TA can improve stock price predictions.

Ayala, García-Torres, Noguera, Gómez-Vela, and Divina (2021) presents a study on enhancing SM index predictions through the integration of machine learning techniques into a TA strategy. The authors suggest a method that optimizes the TIs employed in the strategy by training a machine learning algorithm to determine the most significant indicators. The method was evaluated on several SM indices and found to improve the performance of predictions compared to using TIs alone or a standard machine learning method. The study's findings indicate that combining TA with machine learning can enhance the accuracy of SM index predictions.

Kamara, Chen, and Pan (2022) propose a novel hybrid model for stock price forecasting. The model blends deep learning models and TA to enhance prediction accuracy. The article examines the use of an ensemble technique, which is a combination of multiple models, to improve prediction performance. The model was trained and tested on historical data of a SM index, and the results demonstrate that the proposed model outperforms traditional TA and machine learning models.

Chandar (2022) presents a technique for incorporating TIs in stock trading with a CNN. The author describes the use of the CNN to recognize patterns in TIs such as moving averages and relative strength indexes, and how these patterns can be utilized to make predictions about future stock prices. The study discovered that the CNN model outperformed traditional machine learning models in stock trading by achieving higher prediction accuracy.

Niu, Xu, and Wang (2020) combined variational mode decomposition (VMD) and a LSTM network to predict four stock indices (HIS, FTSE, S&P 500 and IXIC). In this work, authors demonstrate that the combination of two models is much better than using one model. Furthermore, the results of this work show that VMD-LSTM outperform VMD-ELM, VMD-CNN and VMD-BPNN for the SPX and IXIC data series.

Ecer, Ardabili, Band, and Mosavi (2020) compared the predictive performance of two combined models, MLP-GA and MLP-PSO, using two different output functions, Tanh(x) and Gauss. As input data, they mobilized TIs calculated from the historical data of Borsa Istanbul 100 index. The results of this study show that the Tanh(x) output function improves the accuracy of the models and the MLP-PSO model with population size 125 outperform other models utilized in this work.

Orimoloye, Sung, Ma, and Johnson (2020) in their work they compared the abilities of deep feedforward neural networks and shallow architectures for predicting 34 stock price indices in different markets (emerging and developed) and in different time frames (daily, hourly, minute and tick). The results of this paper demonstrate the outperformance of deep NN in different time horizons using ReLu function, except in tick level data.

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Jiang, Liu, Zhang, and Chunyu (2020) combined TIs and macroeconomics variables to predict three major indices in USA (S&P500, NASDAQ 100 and Dow 30). In this paper, they mobilized stacking method to make 20 days predictions and as results they found that this method is outperforms ensemble learning algorithms and deep learning models.

Nikou, Mansourfar, and Bagherzadeh (2019) compared between different machine and deep learning algorithms in predicting the close price of IShares MSCI UK. As result, they found that deep learning models outperform machine learning models.

Goel *et al.* (2022) aimed to predict the close price of Bombay Stock Exchange (BSE) using ANN model and macroeconomics variables. They found that ANN works well in this market and can make accurate predictions with 93%.

From our review of the literature on the prediction of SM indices, two major shortcomings were identified. First, major works are based on predicting developed markets like S&P 500 and NASDAQ 100. Second, a feature selection method to select the pertinent variables has not been undertaken. Our goal is to contribute to filling these significant gaps in the literature.

3. Methodology

This paper aims to predict the SM indices of MASI, MADEX, EGX 30, NASDAQ 100 and S&P 500 using three distinct models: ANN, CNN and LSTM. The data was collected from Investing.com using the Investpy library in Python.

Table 1 displays the periods of the data series, including the observation period, the number of observations and the observations used for training and testing for each index. The observation period spans from a specific date in the past to April 16, 2021, with the number of observations ranging from 4,809 for MASI to 8,962 for NASDAQ 100. The observations used for training data range from 3,847 for MASI to 7,169 for NASDAQ 100, while the observations used for testing data range from 962 for MASI to 1,793 for NASDAQ 100. Deep learning models needs a lot of data to make good predictions, this is why we tried to import the maximum available data for each index and this explains the difference between the number of data in the different indices. The aim of this work is to make predictions of the selected indices and the difference between the number of observations will not impact the results.

We will also compute TIs for each index using the Ta-Lib library, as demonstrated in Table 2. The table lists various TIs used in financial analysis, including their symbols and order, which are commonly employed by traders and investors to analyze historical security or market performance and make predictions about future performance (Ifleh & El Kabbouri, 2022; Ecer *et al.*, 2020).

Our method consists of seven primary steps, as depicted in Figure 1. First, we will extract the data from Investing.com. Second, we will calculate the TIs for each index using the Ta-Lib library. Third, we will use the Correlation Feature Selection method to choose the most relevant features for the model. Fourth, we will split the data into training and testing sets,

Table 1. Number of observations of each index

Index	Period	Number of observations	Train data	Test data
MASI	03/01/2002 to 16/04/2021	4,809	3,847	962
MADEX	03/01/2002 to 16/04/2021	4,807	3,845	962
EGX 30	04/01/1998 to15/04/2021	5,700	4,560	1,140
FTSE 100	03/01/2001 to 16/04/2021	5,125	4,100	1,025
S&P 500	04/01/2006 to 16/04/2021	3,846	3,076	770
NASDAQ 100	26/09/1985 to 16/04/2021	8,962	7,169	1,793

Source(s): Table created by authors

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Technical indicator	Definition	Number of days
Average directional movement index (ADX)	A TI that measures the strength of a trend	14
Average true range (ATR)	A TI that measures the volatility of an asset	14
Bollinger bands (BB)	A TI that measures the volatility of an asset and consists of three lines, a simple moving average, an upper band and a lower band	14; 20
Chaikin A/D line (A/D)	A TI that measures the cumulative flow of money into and out of an asset	
Commodity channel index (CCI)	A TI that measures the deviation of an asset's price from its average price	14
Daily return (Return)	A TI that measures the percentage change in an asset's price from one day to the next	
Directional movement index (DX)	A TI that measures the strength of a trend and the direction of the trend	14
Double Exponential Moving Average (DEMA)	A TI that smooths out price data by taking into account two exponential moving averages	5; 20; 60; 120
Exponential moving average (EMA)	A TI that smooths out price data by taking into account a specified number of past prices	5; 20; 60; 120
Fast stochastic (FastK)	A TI that measures the momentum of an asset's price	
High - Low (High - Low)	A TI that measures the difference between the highest and lowest prices of an asset over a specified period	
Momentum (Mom)	A TI that measures the rate of change in an asset's price over a specified period	10
Money flow index (MFI)	A TI that measures the strength of buying and selling pressure in an asset	14
Moving average convergence divergence (MACD)	A TI that measures the difference between two exponential moving averages	9, 12, 26
On Balance Volume (OBV)	A TI that measures buying and selling pressure by adding or subtracting volume based on whether the price moves up or down	
Open - Close (Open - Close)	A TI that measures the difference between the opening and closing prices of an asset over a specified period	
Percentage price oscillator (PPO)	A TI that measures the difference between two exponential moving averages as a percentage	12, 26
Rate of change (ROC)	A TI that measures the percentage change in an asset's price over a specified period	10
Relative strength index (RSI)	A TI that measures the magnitude of recent price changes to evaluate overbought or oversold conditions	14
Simple moving average (SMA)	A TI that smooths out price data by taking into account a specified number of past prices	5; 20; 60; 120
Standard deviation (STDV)	A TI that measures the variability of an asset's price over a specified period	5
Stochastic (Stoch)	A TI that measures the momentum of an asset's price	
Stop and reverse (SAR)	A TI that identifies potential reversals in an asset's price	
Triangular moving average (TRIMA)	A TI that smooths out price data by taking into account a specified number of past prices and giving more weight to the middle values	5; 20; 60; 120
Triple exponential moving Average (TEMA)	A TI that smooths out price data by taking into account three exponential moving averages	5; 20; 60; 120
Weighted moving average (WMA)	A TI that smooths out price data by taking into account a specified number of past prices and giving more weight to more recent prices	5; 20; 60; 120
Williams' %R (R%)	A TI that measures the momentum of an asset's price and indicates overbought or oversold conditions	14

Source(s): Table created by authors

Table 2. Technical indicators

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

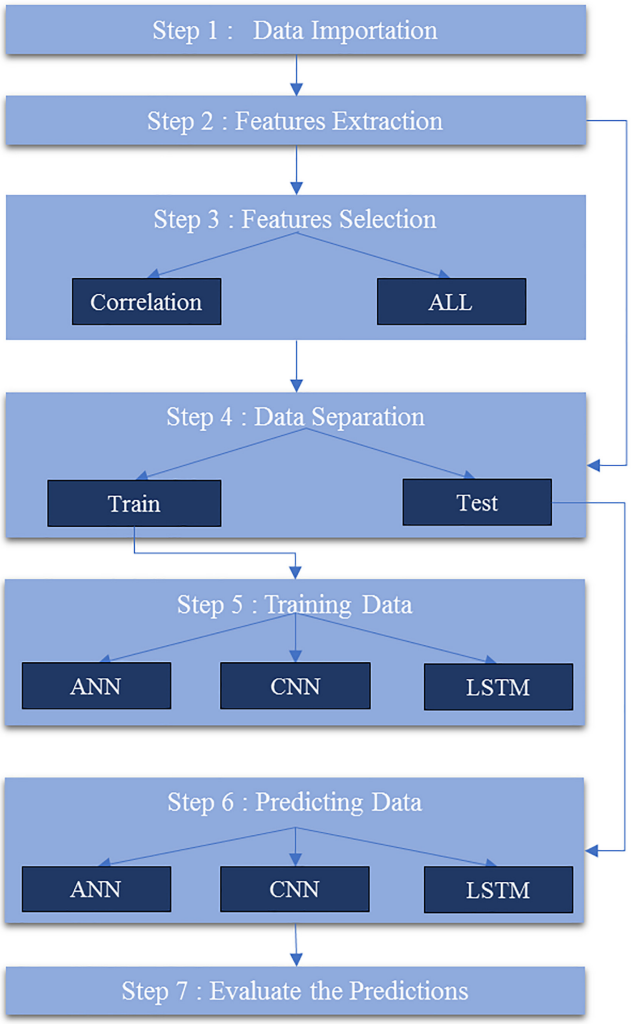


Figure 1.
Proposed method

Source(s): Figure created by authors

with an 80% and 20% ratio, respectively, as indicated in Table 1. Fifth, we will train the data using the above-mentioned models. Sixth, we will make predictions using the trained models. Finally, in the seventh step, we will evaluate the predictions using various accuracy metrics.

3.1 Feature selection

With the rise in data, reducing its dimensionality has become crucial for efficient processing. In various domains, a problem may require a large number of variables, which can cause challenges such as information loss due to noisy data, complexity and extended computation time.

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Feature selection involves extracting a relevant subset of features from the original set. Correlation is a commonly used method of feature selection that measures the degree of connection or the extent to which two variables vary together. The Pearson correlation coefficient is one of the most commonly used measures of correlation.

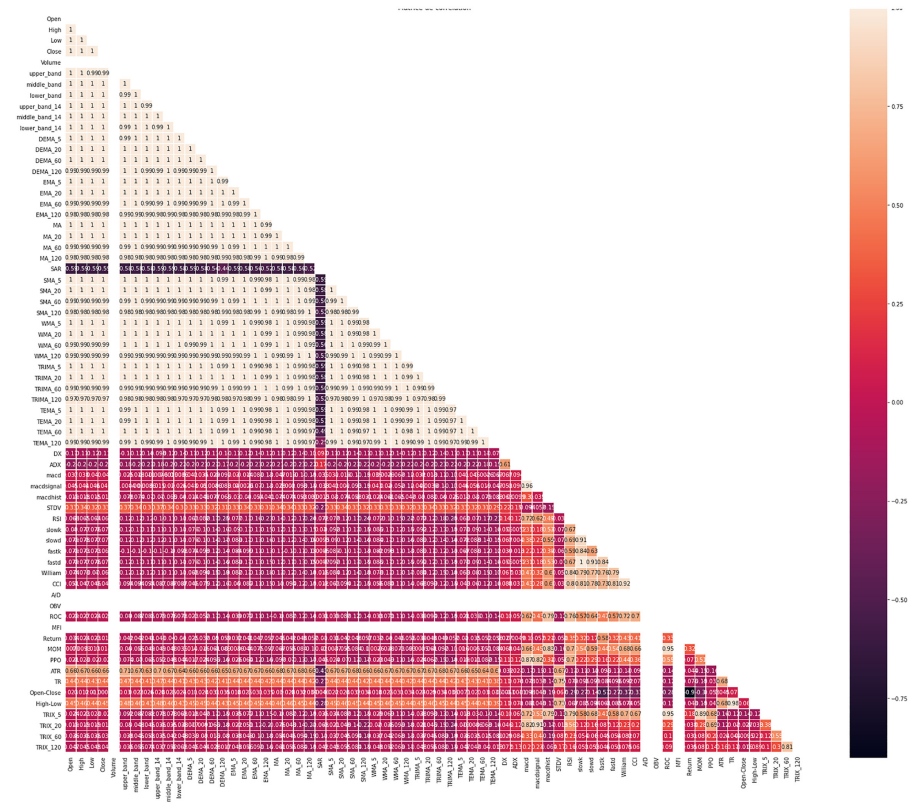
The selection of variables as inputs for a model is based on the principle that they should be correlated with the dependent variable and not correlated with other independent variables.

Using linear correlation as a measure of input quality offers several advantages. Firstly, it enables us to eliminate inputs that are independent of the output. Secondly, it reduces the redundancy among the selected inputs.

Figures 2–7 present the correlation matrix heatmap of the TIs employed in this work in each index. The correlation scales between −1 and 1, where its value is in light (dark) regions it means positive (negative) relationship.

Table 3 presents the relevant variables selected through correlation. The specific indicators included in each index are those that have shown a strong correlation with the close price of that index and independence between them, as determined by the analysis.

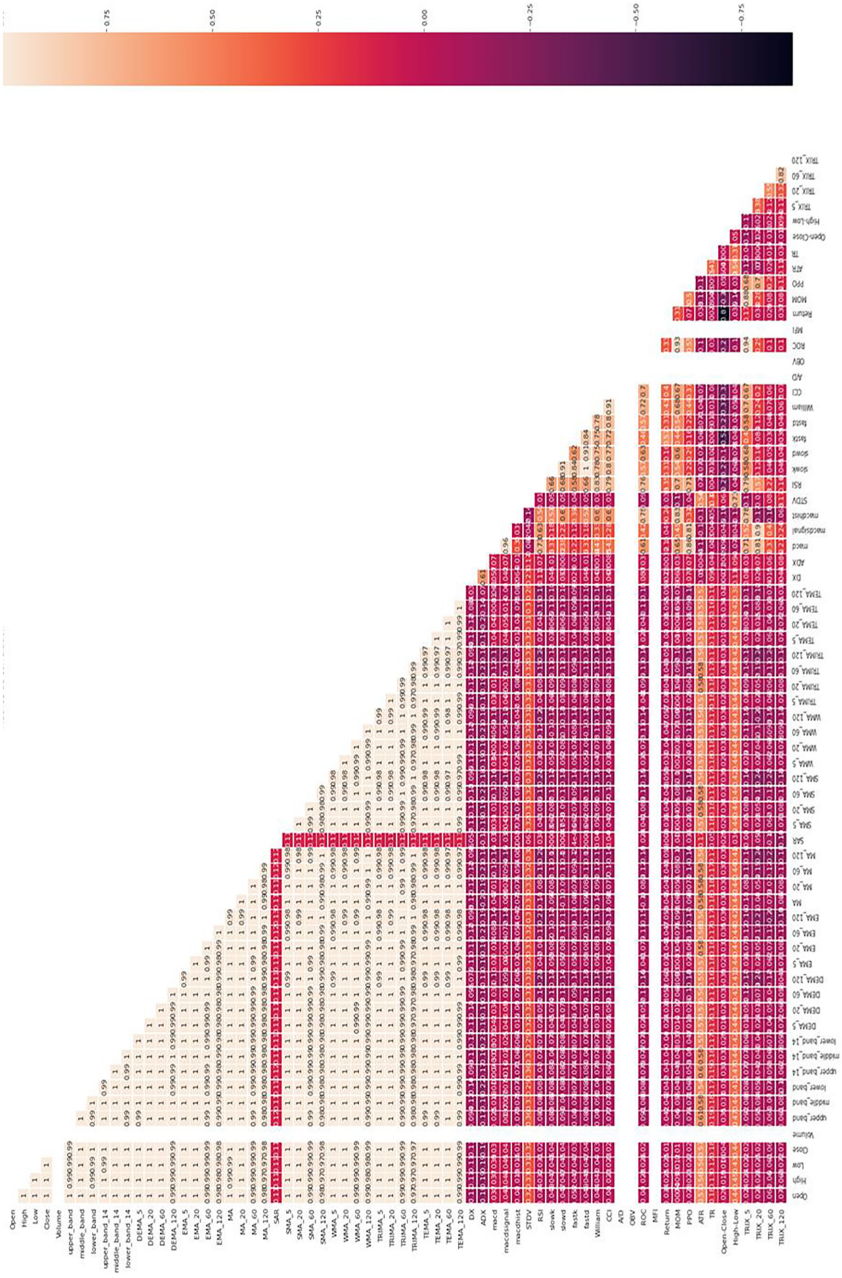
For example, in the MASI index, the indicators SAR, DX, ADX, macd, macdhist, STDV, RSI, Volume, Open-Close and TRIX_60 have been selected as they have a strong correlation with the close price of the index and they are not correlated between them. Similarly, in the



Source(s): Figure created by authors

Figure 2. The correlation matrix heatmap of the variables (MASI)

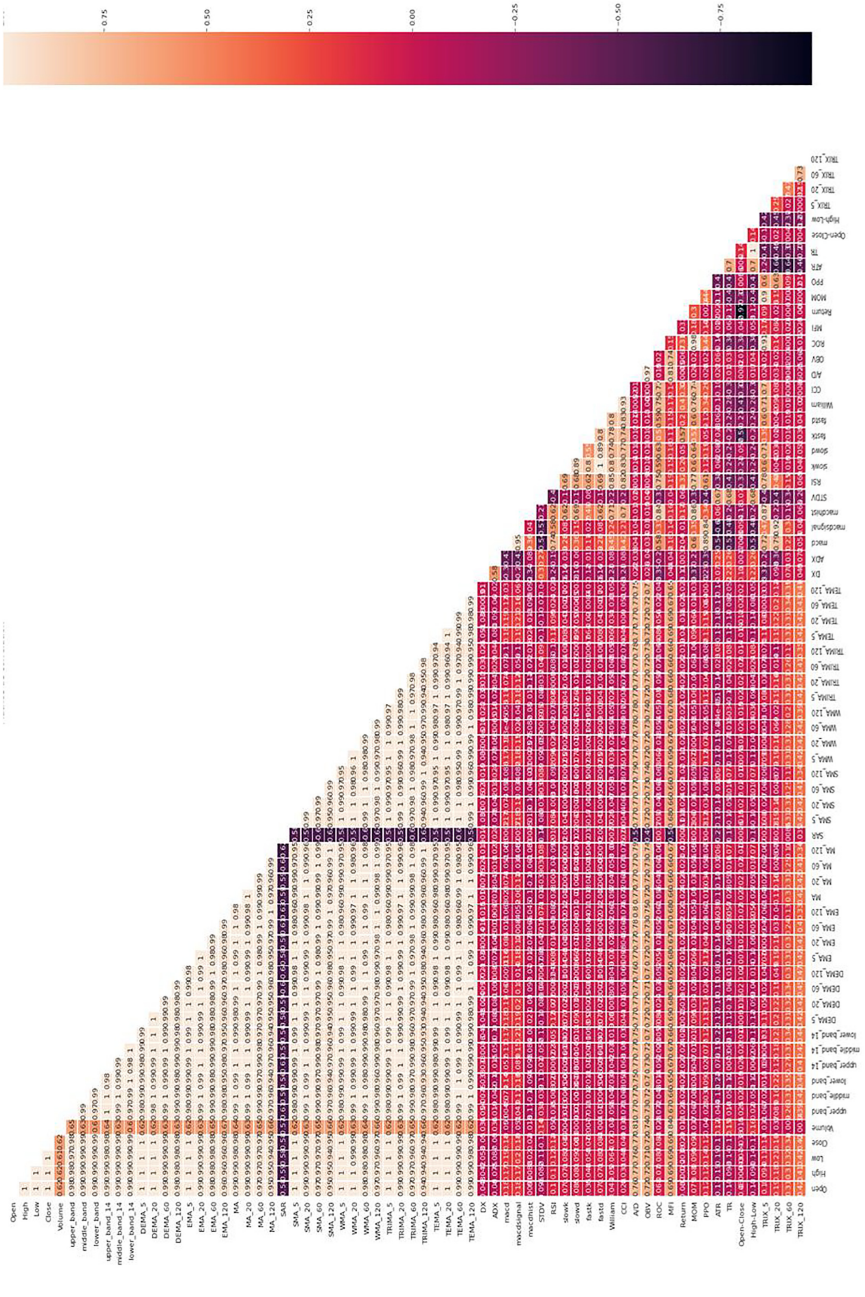
Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026



Source(s): Figure created by authors

Figure 3.
The correlation matrix heatmap of the variables (MADEX)

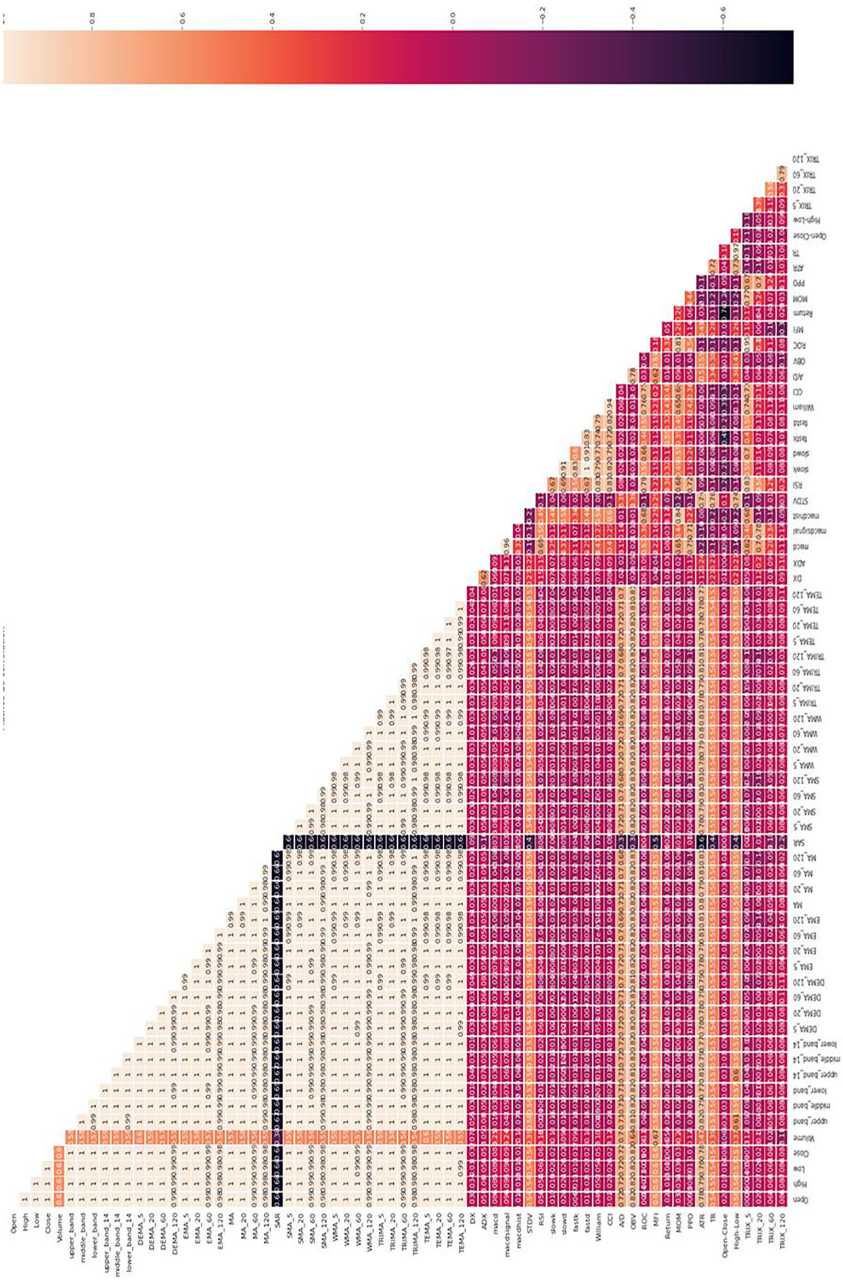
Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026



Source(s): Figure created by authors

Figure 4.
The correlation matrix heatmap of the variables (FTSE 100)

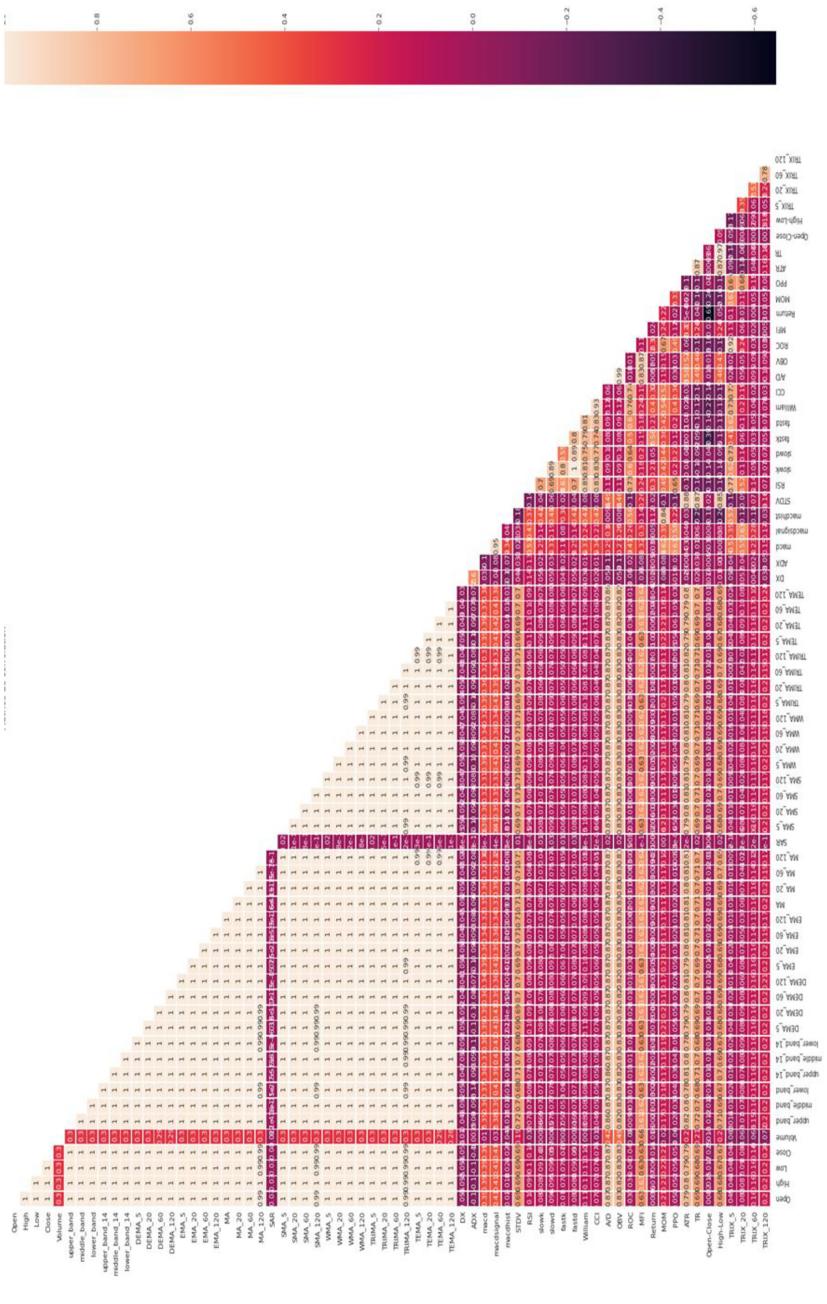
Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026



Source(s): Figure created by authors

Figure 5.
The correlation matrix heatmap of the variables (EGX 30)

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026



Source(s): Figure created by authors

Figure 6.
The correlation matrix heatmap of the variables (NASDAQ 100)

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

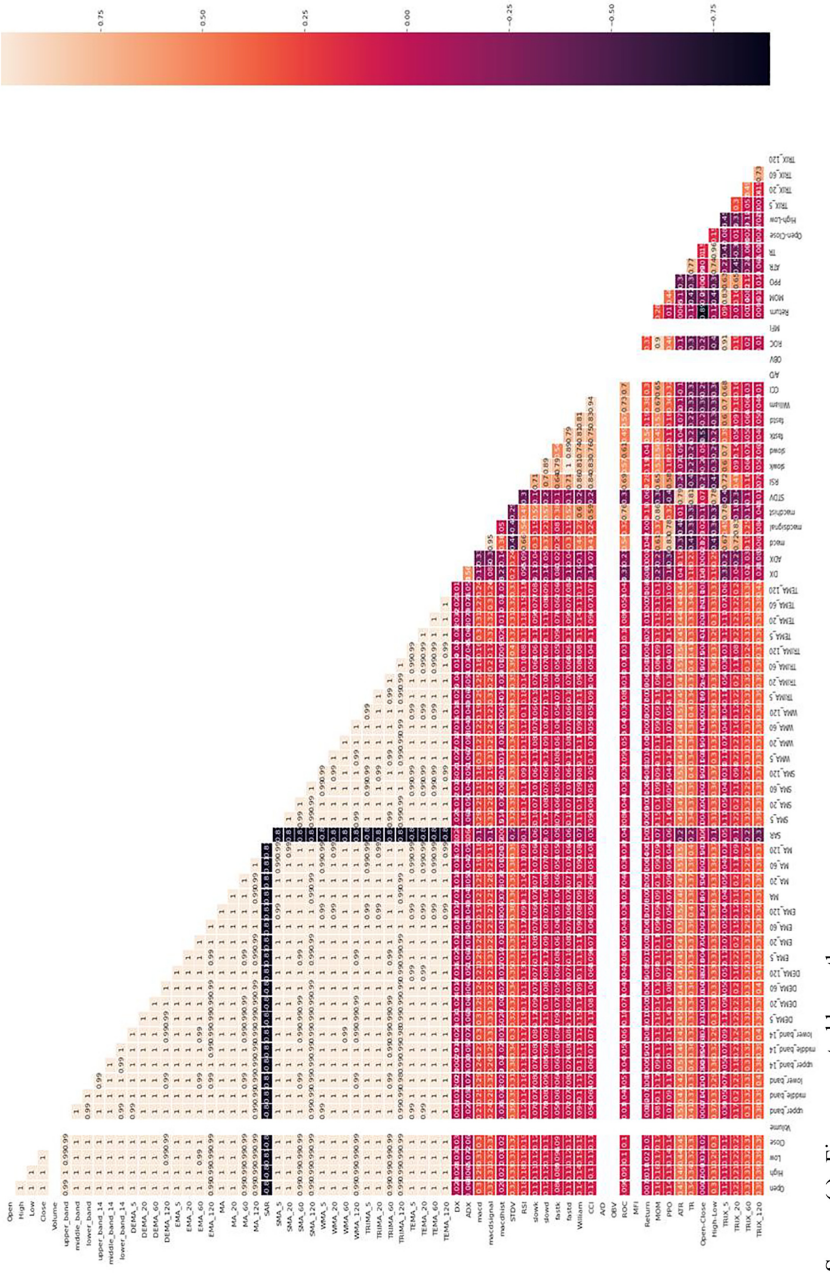


Figure 7.
The correlation matrix heatmap of the variables (S&P 500)

Source(s): Figure created by authors

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

MASI									
SAR	DX	ADX	macd	macdhist	STDV	RSI	Volume	Open-Close	TRIX_60
slowk	A/D	OBV	ROC	MFI	Return	ATR	TR		
MADEX									
SAR	DX	ADX	macd	macdhist	STDV	RSI	Volume	High-Low	Open-Close
slowk	A/D	OBV	ROC	MFI	Return	ATR	TR	TRIX_60	
EGX 30									
SAR	DX	ADX	macdhist	STDV	RSI	slowk	A/D	TRIX_60	Volume
ROC	MFI	Return	PPO	TR	TRIX_20				
FTSE 100									
SAR	DX	macd	macdhist	RSI	slowk	fastk	Open-Close	TRIX_60	Volume
ATR	TR								
NASDAQ 100									
SAR	DX	ADX	macd	macdhist	STDV	RSI	Volume	TRIX_20	TRIX_60
slowk	fastk	ROC	Return	PPO	Open-Close				
S&P 500									
SAR	DX	ADX	macd	macdhist	STDV	RSI	Volume	Open-Close	TRIX_60
slowk	fastk	A/D	OBV	ROC	MFI	Return	ATR		

Source(s): Table created by authors

Table 3. Selected features

MADEX index, the indicators SAR, DX, ADX, macd, macdhist, STDV, RSI, Volume, High-Low and Open-Close have been selected for the same reason.

It's worth mentioning that the specific indicators included in each index may vary, reflecting the application of the correlation analysis to different indexes.

3.2 Mobilised models

In this work we used three models, ANN, CNN and LSTM.

(1) **ANN**

Artificial neural networks (ANNs) are made of multiple units, called perceptrons. Every perceptron simulates the natural neurons of the human brain. Figure 8 shows that the first perceptron receives inputs x, which all are multiplied by weights w. In the following stage, the outcome is derived by comparing it to a threshold, when the weighted sum $\sum_i w_i x_i$ is under the given threshold, then it will be zero, else the output will be one Chopra *et al.* (2019). A multilayer perceptron is called an ANN. ANNs are comprised of several perceptrons organized in layers. The first one gets the inputs and transmits them to the intermediate layers, known as hidden layers. The calculated values in one layer are transmitted to the following layers, where the first one gets the inputs and the last one generates the outputs. In this work, we build our ANN model and train it using 100 epochs and 2 hidden layers because they minimize the loss function.

(2) **CNN**

A convolutional neural network (CNN) is a DL architecture specifically developed for images. CNNs are highly similar to an ordinary NN. They are provided with weights and biases that

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

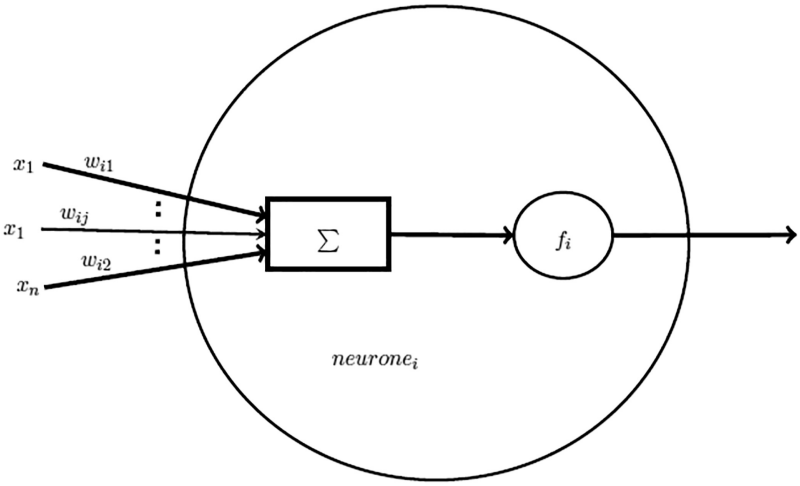


Figure 8.
ANN architecture

Source(s): Figure created by authors

work with the neurons to produce a score that ranks the input data. However, the main difference is that CNNs require that the input data be images, which allows the architecture to be tailored to specific types of data patterns, to be more efficient, and to reduce the number of parameters in the network. For example, since the assumption is that the input data is an image, this allows the network to form associations with only the neighboring pixels instead of the entire image. This avoids unnecessary use of neurons and a variety of parameters. We build our CNN model and train it using 100 epochs and 2 hidden layers because they minimize the loss function.

(3) **LSTM**

In 1997, the LSTM model was first proposed by Hochreiter and Schmidhuber. It removes the problem of gradient vanishing in RNNs. The reason for this problem is that the information is not saved for a long period of time and the gradient in the deepest layers becomes useless.

In order to resolve this problem, the LSTM model includes a memory cell $"C_t"$ which is able to keep the information for a long period of time. Therefore, every memory cell contains three gates, input gate $"I_t"$, forget gate $"f_t"$ and output gate $"O_t"$ (Sethia & Raut, 2019).

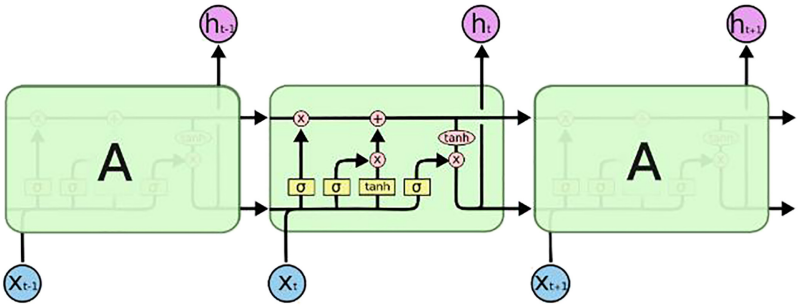
The $"I_t"$ determines whether the input should change the content of the cell, the $"f_t"$ chooses to return the content of the cell to zero and the $"O_t"$ determines whether the content of the cell should offer the output of the neuron.

The Gates are sigmoid functions with a binary value of 0 and 1, where 0 means that nothing passes and 1 that everything passes (See Figure 9). In this work, we build our LSTM model and train it using 100 epochs and 1 hidden layer because they minimize the loss function.

3.3 Forecasting performance measures

There is a wide range of performance measures to judge the precision of the prediction model; in our work we use those measures (Ifleh & El Kabbouri, 2021):

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026



Source(s): Figure created by authors

Figure 9. LSTM architecture

MSLE:

$$\text{MSLE} = \frac{1}{n}\sum_{t=1}^{n}(\log\,(f_t+1) - \log\,(y_t+1))^2. \tag{1}$$

MAE:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - f_t|. \tag{2}$$

MSE:

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t)^2. \tag{3}$$

RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - f_t)^2}. \tag{4}$$

RMSLE:

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\log\,(f_t+1) - \log\,(y_t+1))^2}. \tag{5}$$

where:

$\boldsymbol{y_t}$: Actual value in time period t;

$\boldsymbol{f_t}$: Forecast value in time period t;

$\boldsymbol{n}$: Number of periods forecasted.

The evaluation metrics above give us and idea about the model prediction performance, they compare the predicted close price with reel close price to indicate the accuracy of the model Ecer *et al.* (2020). They should be nearer to zero to offer the better predictions results (Klimberg, Sillup, Boyle, & Tavva, 2010).

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

4. Results and discussion

Table 4 shows the results of the prediction using correlation as a feature selection for different indexes such as MASI, MADEX, FTSE 100, EGX 30, NASDAQ 100 and S&P 500. The table compares the performance of three different machine learning models (ANN, CNN and LSTM) for each index.

The performance of the models is evaluated using different evaluation metrics such as MSE, RMSE, MSLE, RMSLE and MAE. Lower values for these metrics indicate better performance.

For example, in the MASI index, the ANN model has an MSE of 895,324, an RMSE of 29,922, an MSLE of 0.0000, an RMSLE of 0.003 and an MAE of 26,295. This suggests that the ANN model has a relatively lower error rate when compared to other models such as LSTM model, which has an MSE of 14388,787, an RMSE of 119,953, an MSLE of 0.0001, an RMSLE of 0.011 and an MAE of 88,562.

It's worth noting that different models might perform better in different indexes, depending on the complexity and volatility of the SM. Also, different indexes might have different characteristics that might affect the prediction accuracy, so the correlation feature selection should be done carefully and with a good understanding of the SM.

Also,the outcomes show that ANN outperforms other models in predicting MASI, MADEX and FTSE 100 indices. And CNN outperforms other models in predicting EGX 30, NASDAQ 100 and S&P 500.

(1) Comparaison

By comparing the results of predictions using pertinent variables with the predictions using all the variables, we can remark that using all the variables as inputs performs the predictions using ANN model in all markets except in MADEX and NASDAQ 100 where the use of pertinent variables is more interesting.

Also, the predictions using CNN and LSTM combined with selected variables based on correlation outperform the predictions using all the variables in all indices except MASI (See Table 5).

				Correlation		
	Model	MSE	RMSE	MSLE	RMSLE	MAE
MASI	ANN	895.324	29.922	0.0000	0.003	26.295
	CNN	46477.693	215.587	0.0004	0.020	162.880
	LSTM	14388.787	119.953	0.0001	0.011	88.562
MADEX	ANN	2940.619	54.227	0.0000	0.006	34.573
	CNN	29875.191	172.844	0.0004	0.020	101.858
	LSTM	24770.517	157.387	0.0003	0.017	124.712
FTSE 100	ANN	3848.100	62.033	0.0001	0.009	54.151
	CNN	14214.376	119.224	0.0003	0.019	81.102
	LSTM	10170.096	100.847	0.0001	0.009	61.665
EGX 30	ANN	221679.740	470.829	0.0010	0.032	382.507
	CNN	140389.210	374.685	0.0009	0.030	288.876
	LSTM	690774.090	831.128	0.0032	0.057	655.352
NASDAQ 100	ANN	454824.780	674.407	0.0041	0.064	386.738
	CNN	20718.819	143.940	0.0004	0.019	93.993
	LSTM	1787310.500	1336.903	0.0198	0.141	799.798
S&P 500	ANN	196041.980	442.766	0.0183	0.135	281.361
	CNN	7620.777	87.297	0.0009	0.030	68.823
	LSTM	43336.416	208.174	0.0036	0.060	141.945

Source(s): Table created by authors

Table 4. Accuracy metrics of models using correlation-based features selection

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Table 5. Comparison between accuracy metrics of models using correlation-based features selection and models using all features

		Correlation					Without selection				
	Model	MSE	RMSE	MSLE	RMSLE	MAE	MSE	RMSE	MSLE	RMSLE	MAE
MASI	ANN	895.324	29.922	0.0000	0.003	26.295	779.053	27.912	0.0000	0.003	20.376
	CNN	46477.693	215.587	0.0004	0.020	162.880	45103.495	212.376	0.0004	0.020	125.739
	LSTM	14388.787	119.953	0.0001	0.011	88.562	10877.065	104.293	0.0001	0.010	65.768
MADEX	ANN	2940.619	54.227	0.0000	0.006	34.573	3691.069	60.754	0.0000	0.007	39.180
	CNN	29875.191	172.844	0.0004	0.020	101.858	110435.830	332.319	0.0015	0.039	240.901
	LSTM	24770.517	157.387	0.0003	0.017	124.712	103689.530	322.009	0.0014	0.037	260.761
FTSE 100	ANN	3848.100	62.033	0.0001	0.009	54.151	2174.872	46.636	0.0000	0.007	40.362
	CNN	14214.376	119.224	0.0003	0.019	81.102	20074.403	141.684	0.0005	0.022	105.591
	LSTM	10170.096	100.847	0.0001	0.009	61.665	14860.848	121.905	0.0003	0.018	96.425
EGX 30	ANN	221679.740	470.829	0.0010	0.032	382.507	72781.951	269.781	0.0003	0.018	221.988
	CNN	140389.210	374.685	0.0009	0.030	288.876	889860.450	943.324	0.0051	0.071	837.910
	LSTM	690774.090	831.128	0.0032	0.057	655.352	1122112.800	1059.298	0.0055	0.074	869.314
NASDAQ 100	ANN	454824.780	674.407	0.0041	0.064	386.738	423427.350	650.713	0.0037	0.061	363.367
	CNN	20718.819	143.940	0.0004	0.019	93.993	527106.150	726.021	0.0058	0.076	460.892
	LSTM	1787310.500	1336.903	0.0198	0.141	799.798	4840166.800	2200.038	0.0709	0.266	1450.848
S&P 500	ANN	196041.980	442.766	0.0183	0.135	281.361	4738.519	68.837	0.0004	0.020	50.601
	CNN	7620.777	87.297	0.0009	0.030	68.823	34252.489	185.074	0.0040	0.063	140.560
	LSTM	43336.416	208.174	0.0036	0.060	141.945	77477.782	278.348	0.0067	0.082	187.513

Source(s): Table created by authors

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

In other words, it's better to predict MADEX and NASDAQ 100 combining selected variables with ANN and CNN, respectively. Compared to other researchers Kamara *et al.* (2022), Sang and Di Pierro (2019), Chandar (2022), Ifleh and El Kabbouri (2021, 2022), in this work we employed new methodology to make predictions and we worked on different markets (emerging and developed). The results show the markets where we can employ correlation feature selection.

5. Conclusion

In this study, we aimed to predict six different SM indices using various machine learning models (ANN, CNN and LSTM). We also examined whether using variables selected based on correlation would result in more accurate predictions than using all variables.

We build and train our models using 100 epochs and 2 hidden layers, except in LSTM model we use 1 hidden layer, because they minimize the loss function.

Our results showed that ANN outperformed other models in predicting MASI, MADEX and FTSE 100 indices, while CNN outperformed other models in predicting EGX 30, NASDAQ 100 and S&P 500.

When comparing the results with predictions made using all features, we found that the combination of ANN and all variables generally provided better results, except in the case of MADEX and NASDAQ 100. Additionally, predictions made using CNN and LSTM combined with selected variables based on correlation outperformed predictions made using all variables for all indices except MASI.

There are many possibilities for improving the predictive ability of this study. One promising avenue is to explore alternative feature selection models, such as the random forest algorithm, which could yield more robust results. In addition, combining various predictive models together appears to be another viable strategy.

Broadening the scope by integrating a wider range of features could prove useful in highlighting the most important elements of the predictions. These refinements could then inform and guide future research efforts.

In addition, we recommend that researchers consider incorporating additional variables, such as macroeconomic and sentiment indicators, into their research. Joining the results of these variables to our findings, valuable information can be gained to enrich the body of knowledge.

References

Agrawal, M., Khan, A. U., & Shukla, P. K. (2019). Stock price prediction using technical indicators: A predictive model using optimal deep learning. *International Journal of Recent Technology and Engineering (IJRTE) ISSN*, *8*(2), 2277–3878.

Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems*, *225*, 107119.

Chandar, S. K. (2022). Convolutional neural network for stock trading using technical indicators. *Automated Software Engineering*, *29*(1), 1–14.

Chopra, S., Yadav, D., & Chopra, A. N. (2019). "Artificial neural networks based Indian stock market price prediction: Before and after demonetization". *International Journal of Swarm Intelligence and Evolutionary Computation*, *8*(1). doi: 10.4172/2090-4908.1000174.

Dash, R. & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science*, *2*(1). doi: 10.1016/j.jfds.2016.03.002.

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Ecer, F., Ardabili, S., Band, S. S., & Mosavi, A. (2020). Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction. *Entropy*, *22*(11), 1239. doi: 10.3390/e22111239.

Efat, I. A., Bashar, R., Imtiaz Ud-Din, K. M., & Bhuiyan, T. (2018). Trend estimation of stock market: An intelligent decision system. In *International Conference on Cyber Security and Computer Science (ICONCS'18)*, Safranbolu, Turkey, Oct 18-20, 2018.

Fama, E. F. (1965). Random walk in stock market prices. *Financial Analysts Journal*, *21*, 55–59.

Goel, H. & Singh, N. P. (2022). Dynamic prediction of Indian stock market: An artificial neural network approach. *International Journal of Ethics and Systems*, *38*(1), 35–46. doi: 10.1108/IJOES-11-2020-0184.

Ifleh, A. & El Kabbouri, M. (2021). Moroccan stock market prediction using LSTM model on a daily data. In A. Sheth, A. Sinhal, A. Shrivastava, & A. K. Pandey (Eds.), *Intelligent Systems. Algorithms for Intelligent Systems*. Singapore: Springer.

Ifleh, A. & El Kabbouri, M. (2022). Prediction of Moroccan stock price based on machine learning algorithms. In A. Abraham, N. Gandhi, T.Hanne, T. P. Hong, T. Nogueira Rios, & W. Ding (Eds.), *Intelligent Systems Design and Applications. ISDA 2021. Lecture Notes in Networks and Systems* (Vol. 418). Cham: Springer.

Jiang, M., Liu, J., Zhang, L., & Chunyu, L. (2020). An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*, *541*, 122272. doi: 10.1016/j.physa.2019.122272.

Kamara, A. F., Chen, E., & Pan, Z. (2022). An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Information Sciences*, *594*, 1–19.

Klimberg, R. K., Sillup, G. P., Boyle, K. and Tavva, V. (2010). Forecasting Performance Measures—what are their practical meaning? In K. D. Lawrence & R. K. Klimberg (Eds.), *Advances in Business and Management Forecasting* (Vol. 7, pp. 137-147). Bingley: Emerald Group Publishing. doi: 10.1108/S1477-4070(2010)0000007012.

Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2019). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock. *Applied Soft Computing Journal Exchange Market*, *91*.

Naik, N. & Mohan, B. (2019). Optimal feature selection of technical indicator and stock prediction using machine learning technique. *ICETCE 2019, CCIS*, *985*, 261–268.

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting Finance and Management*, *26*, 164–174. doi: 10.1002/isaf.1459.

Niu, H., Xu, K., & Wang, W. (2020). A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network. *ApplIntell*, *50*, 4296–4309. doi: 10.1007/s10489-020-01814-0.

Orimoloye, L. O., Sung, M.-C., Ma, T., & Johnson, J. E. V. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systemswith Applications*, *139*, 112828. doi: 10.1016/j.eswa.2019.112828.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2014). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*, 2162–2172.

Qiu, M. & Song, Yu (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS One*, *11*(5), e0155133. doi: 10.1371/journal.pone.0155133.

Ratto, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60–70.

Sahoo, S. & Mohanty, M. N. (2020). "Stock market price prediction employing artificial neural network optimized by Gray Wolf optimization". *Advances in Intelligent Systems and Computing*, *1030*, 77–87.

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026

Sang, C. & Di Pierro, M. (2019). Improving trading technical analysis with tensorflow long short-term memory (LSTM) neural network. *The Journal of Finance and Data Science*, *5*(1), 1–11.

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation*.

Sethia, A. & Raut, P. (2019). Application of LSTM, GRU & ICA for stock price prediction. In *Proceedings of ICTIS 2018* (Vol. 2). doi: 10.1007/978-981-13-1747-7_46.

Sezer, O. B., Ozbayoglu, M., & Dogdu, E. (2017). A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters. In *Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems*, Chicago, Illinois, USA, CAS October 30 – November 1.

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2019). "Financial time series forecasting with deep learning: A systematic literature review 2005-2019". Papers 1911.13288, arXiv.org. doi: 10.1016/j.asoc.2020.106181.

Further reading

Chang, P.-C., Liao, T. W., Lin, J.-J., & Fan, C.-Y. (2011). A dynamic threshold decision system for stock trading signal detection. *Applied Soft Computing*, *11*. doi:10.1016/j.asoc.2011.02.029.

Ou, P. & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science*, *3*, 28.

Yodele Adebiyi, A., Ayo Charles, K., Adebiyi Marion, O., & Otokiti Sunday, O. (2012). Stock Price prediction using neural network with hybridized market indicators. *Journal of Emerging Trends in Computing and Information Sciences*, *3*(1), 1–9.

Corresponding author
Abdelhadi Ifleh can be contacted at: iflehabdo@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Downloaded from ftp.nowpublishers.com/agjsr/article-pdf/42/4/1237/9516345/agjsr-02-2023-0070.pdf by guest on 15 October 2026