International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Gene Expression Data Classification Using Machine Learning with SigFeature: A Novel Significant Feature Selection Method

[1]Dr.Yogesh Suresh Deshmukh, [2]Prasad Raghunath Mutkule, [3]Dr. Rajesh Kedarnath Navandar, [4]Dr. Pranoti Prashant Mane,[5]Dr. Shubhangi Milind Joshi,[6]Dr. Madhuri Pravin Borawake, Govind M. Poddar[7]

Submitted: 27/12/2023 **Revised**: 03/02/2024 **Accepted**: 11/02/2024

Abstract: In current scenario, machine learning has drawn increasing interest in stock market analysis. To analyse the stock market data for identifying significant feature, data mining and machine learning techniques can be applied. Identifying domain specific feature is a continuous, iterative and logical process. Feature recognition has wide range of applications but no single approach focusses on stock market analysis. Wavelets have its own advantage in tremendous applications however remains less explored in the field of economics and finance. This proposed methodology analysed the new feature, trading interval, for prediction of stock prices which is complex, challenging and need tremendous efforts. This system uses wavelets for identifying domain specific feature in stock market data. General forecasting models were used to forecast the denoised signals. Further the forecast model is selected based upon the performance measure, coefficient of determination with high values. The selected model is used to forecast the share prices. Then performance measures such as RMSE, MAE, MAPE and Theil U were calculated for which trading length consider for analysis. The optimal trading length was found out based on the lowest values of performance measures. Finally, purchase decision making rules were applied to evaluate the accuracy of the selected model and based on that the recommendation of buy or sell was given. This methodology has three proposed approaches; the first approach identifies and removes noise from the stock data efficiently. The second approach involves feature recognition using wavelets on stock market data and the third approach concentrates on analysing stock data, which identifies new feature for economic and financial applications. Finally, to assist investors in making stock market decision, a decision support system with trading interval is presented.

Keywords: *Data Mining, Feature Recognition, Forecasting Models, Machine Learning, Stock Market Analysis, Wavelets, RMSE, MAE, MAPE, Theil U, Purchase Decision Making.*

1. Introduction

In various fields including trading, finance, statistics, and computer science, the direction of stock price movements and prediction of its future values are extensively studied. As a result of the continuous thirst of stock market traders for finding a better method of predicting the direction and movement of stock prices, this model is naturally the result of the desire to find an improved approach in order to buy and sell stocks in profits at the right time [1-2].

Stock analysis and investment decisions are usually based on fundamental or technical analysis. An approach based on fundamental analysis involves analysing an economy, an industry, and a company. Alternatively, you can study historical price patterns to use technical analysis. Assuming a correlation exists between current price and historical price is one of the fundamentals of technical analysis. Thus, the analyst can recommend when to enter and exit the market for a particular company.

In security analysis the valuation of stock is a predominant criterion and analyst used various statistical models such as correlation, regression, charts and technical indices [3]. The traditional techniques were slowly becoming supplementary for analysts. So, the dominant strategy adopted by analysts in recommending buying and selling of shares, entry and exit of market were heavily relying upon improved algorithms.

1.1. Traditional Techniques

Normally two classes of tools are used by investors and traders to make share purchase decision. They are fundamental and technical analysis. The analysis tool

[1] Assistant Professor, Department of Information Technology, Sanjivani College of Engineering, Kopargaon-423603
yogeshdeshmukhit@sanjivani.org.in
[2] Assistant Professor, Department of Information Technology, Sanjivani College of Engineering, Kopargaon-423603
Mutkuleprasad000@gmail.com
[3] Associate Professor, Department of Electronic & Telecommunication Engineering, JSPM Jayawantrao Sawant College of Engineering Hadaspar,Pune , India
navandarajesh@gmail.com4
[4]PhD (Electronics and telecommunication engineering), Associate Professor and HOD,MES's Wadia College of Engineering, Pune
ppranotimane@gmail.com
[5] PhD (Electrical & Electronics Engineering), Associate, Department of Electronics and Communication Engineering , School of Engineering and Sciences Loni Kalbhor, Pune.
sjoshi276@gmail.com
[6]AP and HOD ,Department of Computer Engineering, PDEA'S College of Engineering Manjari Bk, Pune
madhuri.borawake@gmail.com
[7]Associate Professor, NES's Gangamai College of Engineering. Nagaon, Dhule (Maharashtra), India
gmpoddar2811@gmail.com

basically aimed at analysing and predicting shifts in supply and demand. The shifts in supply and demand, usually acts as the basis for economic and fundamental forecasting [4]. If the number of buyers for a stock are more than the sellers the condition implies more demand for the stock so as per the economic theory the price of the stock will raise. If the number of sellers are more than number of buyers it indicated more supply in the market thus the stock's prices will fall. Based on the ability of the investor to foresee the condition, they will get profit and also knows when to enter or exit the market. The fundamental analysis based upon economic, industry and company analysis [5-7]. In economic study the macroeconomic factors such as Gross Domestic Product, measures of consumer confidence, interest rate, inflation rate, money and stock price, Government policies with special reference to fiscal policy and monetary policy shows the direction of economy. This is helping the investors take a decision on potential to invest. In barometric approach using economic factors such as leading, lagging and coincidental factors the investors get knowledge on how the various sectors of economy are going to perform. Based on this the investors can select the type of industrial sectors that are going to perform better and which are going to become worse. An analyst must analyze various sectors of the economy in terms of industries after conducting an industry analysis and determining the way the economy will evolve in short-, medium- and long-term. In point of view of investment the industries can be classified based on its position in industrial life cycle [8]. The industry life cycle classified the industries into pioneer industry, growth industry, defensive industry and declining industry. The classification can be done by various techniques such as analysis of state of competition, cost and profitability and technology and research. In order to estimate a company's intrinsic value, basic financial variables are considered [9]. In addition to these variables, there are many other variables such as sales, profit margins, tax rates, depreciation rates, asset utilization and sources of funding. All of these variables must be taken into account when estimating the company's intrinsic value. Analyzing these variables can help investors determine if a company is undervalued or overvalued [10]. This also helps investors make better decisions when investing in a company. The dividend discount model can be used to value common stocks. For short-term estimate of intrinsic value, one can use an earnings multiplier model. Earnings per share (EPS) are divided by multiples or P/E ratios to determine intrinsic value.

This method helps investors to understand the company's performance compared to its peers and the industry as a whole. It also allows them to determine whether the company is overvalued or undervalued in the current market [12-13]. Analyzing a company's financial reports such as its income statement, balance sheet, and cash flow statement allows investors to compare the company's performance to that of its peers and the industry average. This allows investors to identify any discrepancies and make informed decisions about whether the company is a good investment. In the process of company analysis, major financial information is gleaned from the company's financial statements. Financial statements include the following.

i) Statement of Balances
ii) Financial Statements
iii) An analysis of cash flow

By analyzing past market data and analyzing price and volume, technical analysis forecasts the direction of prices. Technical analysis was extensively incorporated into behavioral economics and quantitative analysis. There are many techniques used in technical analysis. The most popular analyses were candle stick charting, Dows theory and Elliot wave theory. By using the indicators, analysts hope to gain insight into the likely future direction of the market by analyzing past price and volume data. The goal is to identify patterns or trends in the market data that can be used to make predictions about the future. An asset's trend, and the likelihood that it will continue, can be determined by these indicators. Indicators like relative strength indexes and MACD are also looked at by technicians. In addition to options, you can look at put/call ratios and changes in implied volatility. By studying these indicators, technicians can gain an understanding of an asset's potential trend and how likely it is to continue. Consequently, put/call ratios, implied volatility, and other data points can be analyzed to develop a more informed trading strategy [14-16].

EMH translates into random walk theory as efficient market hypothesis. Currently available information about the company's value is fully reflected in current stock prices; therefore, there can be no excess returns from it profits, by using this information. It appears very difficult and unlikely to profit from predicting price movements. An investment environment was compared with the EMH theory. When prices adapt quickly and without bias to new information, a market is said to be efficient.

1.2. Modern Techniques

The evidences for contradictions of EMH theory, led to the niche finding in forecasting the stock prices in Indian Stock Markets. This concept led to the use of the regression and time series techniques to predict the price movements and profiting in stock market. Technical indicators permitted the exploitation of tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, and low-cost solutions. In spite of all these it was found very difficult to accurately predict the future stock prices. This was due to the fact that stock market trends tend to behave in a nonlinear, uncertain, and non- stationary. This situation led to the fact that failure of arriving at a consensus among experts on the effectiveness forecasting financial time series. This situation necessitated the search for finding more techniques that included artificial neural network, Auto regressive conditional hetroskedasticity (ARCH), Generalized Auto regressive conditional hetroskedasticity (GARCH) and Artificial bee colony algorithm (ABC). These are found to be the modern techniques in practice for analysing time series effectively. The failure of GARCH in addressing non stationary trends of stock market movements led to the use of wavelet analysis. Wavelet analysis is a relatively new eld in signal processing. Wavelets are mathematical functions that decompose data into different frequency components, after which each component is studied with a resolution matched to its scale, where a scale denotes a time horizon. Wavelet ltering is closely related to the volatile and time varying characteristics of the real-world time series and is not limited by the stationarity assumption [17].

2. Review of Literature

Cheng et al. [1] the study delves into the realm of stock price prediction by examining and contrasting two prominent approaches: Fundamental Analysis and Technical Analysis. Fundamental Analysis is elaborated upon in terms of key performance indicators such as Return on Equity (ROE), Earnings per Share (EPS), and Price-Earnings (PE) ratio, explaining their role in predicting stock prices. In contrast, Technical Analysis is discussed with a focus on the evaluation of historic stock price entailing the usage of graphs, technical pointers, and prediction models to decode trends over a specific timeframe. The ultimate objective of technical analysts, which is to construct a precise model forecasting the dynamics of stock price, is also explicitly covered. The paper serves as a thorough exploration of these methodologies in the context of stock price prediction, aiming to equip market players with a robust understanding and comparison of these approaches.

The paper, authored by Chen et al. [2], undertakes an exhaustive analysis and review of various fundamental analysis models—specifically, the Discounted Dividend Model (DDM), Multiplier Models and the Discounted Cash Flow Model (DCFM)—for the purpose of developing an improved stock valuation model. After examining these models under conditions of optimal financial market efficiency, the research was directed towards the Residual Income Model (RIM) or the Ohlson Model. The findings suggest the RIM to be the most effective and credible model for predicting stock values, boasting applicability in both emerging and developed markets without necessitating market financial efficiency. This suggests a significant potential for its broad-based adoption in stock valuation processes. Its superior accuracy and feasibility highlight the Residual Income Model's prominence as a preferred tool in the field of stock valuation.

In the paper [3], Chrsitopher M et.al. investigates the critical role of fundamental research in mitigating investment risk and enhancing return on investment. This work establishes that investment decisions in long-term capital market instruments are significantly hinged on fundamental factors. The performance of these securities is not solely dependent on the company's performance, but also on industrial and economic dynamics. Therefore, the paper suggests a more comprehensive approach to investment decisions that factors in industrial and economic elements for an optimal financial outcome.

The paper [4] presents a comprehensive technical analysis of select companies within the CNX Nifty, focusing on three banking institutions chosen out of a pool of over 9000 listed entities. Implementing a simple random sampling methodology, secondary data was collected from spider software to conduct the technical analysis. Using candlestick charts and pertinent indicators as principal tools, the study sought to identify high-potential stocks for investment. The price trend of fifteen stocks was meticulously examined using MACD (Moving Average Convergence Divergence) and RSI (Relative Strength Index) charting techniques. The findings underscore the substantial role that technical analysis plays in assisting investors to make judicious investment decisions.

Nitim et.al. [5] the study compared the stock prices of the companies using the Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) technical indicators to determine whether these indicators could be used to predict the stock prices accurately. These measures of technical analysis help to identify trends in stocks and provide clues as to when it might be a good time to buy or sell. They can also show when the stock market is overbought or oversold and indicate when investors should be more cautious. The ARIMA model is a statistical method used to analyze time-series data. It takes into account the past values of a stock index, as well as any seasonal trends, and uses this information to make predictions about future values. This model is particularly useful for short-term predictions because it can take into account changes in the market quickly, giving financial specialists a better idea of when to buy or sell stocks.

2.1. Feature Recognition

The definition of a “feature” can be broad and entirely application dependent. Feature is an attribute or interesting property which describes an application. The aim of feature recognition (FR) is algorithmically extracting higher level entities from lower level elements of an application. Even in today's deep learning technologies the recognition of significant feature in trading system is fiendishly hard due to daily up and downs of the market. In trading feature recognition can be used to identify profit/lost in trading system. For generating high returns in automated trading strategies feature recognition is inevitable. A dedicated set of computational methods is required for feature recognition, since it faces a variety of unique challenges. In terms of system design and implementation, feature recognition offers a greater degree of freedom. First of all, feature recognition doesn't have a consistent structure, so you don't know what to do (what features have to be recognized, what's significant, etc.). There are also several factors to consider in feature recognition. It may not be appropriate or feasible to apply FR techniques to new recognition problems that were successful in previous problems. Additionally, feature recognition algorithms require relatively high computational power, and it can be a challenge to identify features that have the highest recognition accuracy. Furthermore, feature recognition algorithms are also sensitive to noise and can be difficult to scale up.

2.2. Wavelets

"The wavelet transform is a tool that cuts up data, functions or operators into different frequency components, and then studies each component with a resolution matched to its scale. Shifting refers to the process of translation of signal (t) to (t-s) where s is the shifting parameter. By applying multiple shifts, we can get multiple translations of the same mother wavelet. This is called wavelet packet and the wavelet packet representation of the signal (t) is given in Equation 1. The scaling and shifting of the signal (t) can be combined to form wavelet packet transform which is used to represent the signal in terms of shifted and scaled version of the mother wavelet.

$$F_{eq} = C_f / S\partial t \tag{1}$$

where $C_f = Centre\ Frequency$
$\partial t = Sampling\ Interval$

Low frequency signals are better represented with wider windows and high frequency signals are better represented with shorter windows. This gives the user the ability to choose the appropriate scale factor in order to achieve the desired resolution. In contrast to compressed wavelets, stretched wavelets capture signals that vary slowly over time. Stretched wavelets are especially useful for analyzing signals with long-term trends. They can also be used to identify slow changes in signals that are otherwise overlooked by standard techniques. Wavelet shifted and centered at k is represented by constant. Wavelets are shifted and scaled according to the application to identify features in the signal. The following figure shows how wavelets are scaled and shifted.

2.3. Continuous Wavelet Transform (CWT)

Time frequency analysis of data can be performed using wavelets. By using mother wavelets, which are translated and dilated versions of a basis function, wavelet transformation analyzes a signal in time domain. Wavelet transformation decomposes signals into frequency components and is useful for analyzing non-stationary signals. The components can then be reconstructed to obtain the original signal. The use of wavelet transformation for time frequency analysis is becoming increasingly popular in many fields. Continuous wavelets can be used to analyze and filter localized frequency components over a period of time. Wavelets can also be used for signal compression, denoising, and feature extraction. They have proven to be versatile tools for signal processing and offer improvements over traditional Fourier Transforms. In CWT, as far as signal reconstruction is concerned, the information provided by CWT coefficient is highly redundant and needs more resources and computation time. This work aimed to CWT for identifying interval- based analysis of stock data in future. The CWT is defined in Equation 2

$$CWT_x^{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_t^0 x(t) \psi * \frac{(t-b)}{a}\, dt \tag{2}$$

where
$CWT_x^{\psi} = Continuos\ Wavelet\ Transform\ of\ Signal\ x(t)\ using\ wavelet$

a=Scale parameter
b= Translation Parameter
$\frac{1}{\sqrt{a}}$ = A Normalization Constant
$x(t)$ = Signal to be analyzed
t, b = Shifting and Scaling

2.4. Discrete Wavelet Transform (DWT)

Continuous and discrete wavelet transform differs by the manner in which the wavelet is shifted and scaled. DWT is easy to implement when compared with CWT. It is very difficult to separate the noise by looking into the time domain representation. Denoising and compressing signals are the main applications of discrete wavelets. To denoise, the largest coefficients have to be kept. The signal is decomposed into two parts: low-frequency and high-frequency components. The low-frequency components contain most of the information while the high-frequency components contain the noise. The noise can then be removed by discarding the high-frequency components. By projecting signals (S(t)) onto wavelet basis, any signal (S(t)) can be decomposed [25].

2.5. Haar Wavelet Transform (HWT)

Haar wavelet is the oldest and simplest of wavelets and captures fluctuations between adjacent observations. It resembles a square form. The wavelet function, scaling function of Haar wavelet (ψ(t)), is defined in Equation 3

$$\psi(t): \begin{cases} 1 & for\ 0 < t < 1/2 \\ -1 & for\ 1/2 < t < 1 \\ 0 & otherwise \end{cases} \tag{3}$$

3. Scenario of Stock Exchanges

Exchange-traded funds (ETFs) are a type of security that tracks an index, commodity, or basket of assets like an index fund, but trades like a stock on an exchange. The listed companies must adhere to the regulations of the exchange, including filing financial statements and adhering to corporate governance standards. The New York Stock Exchange (NYSE) was the largest stock exchange operator worldwide, with a market capitalization of listed companies of around 33.1 trillion U.S. dollars. The NASDAQ followed in second place with a market capitalization of around 12.7 trillion U.S. dollars. Tokyo Stock Exchange ranked third, with a market capitalization of around 6.7 trillion U.S. dollars. All given in Table 1 below.

Table 1. Stock Exchanges in the World

Largest Stock Exchanges of the world	Market Capitalization (In Trillion U.S. Dollars)
NYSE, UNITED STATES	23.14
NASDAQ, UNITED STATES	10.38
JAPAN EXCHANGE GROUP	6.29
SHANGHAI STOCK EXCHANGE , CHINA	5.02
LONDON STOCK EXCHANGE, LONDON	4.6
HONG KONG STOCK EXCHANGE, HONG KONG	4.44
SHENZHEN STOCK EXCHANGE , CHINA	3.55
DEUTSCHE BÖRSE AG, GERMANY	2.34
BOMBAY STOCK EXCHANGE, INDIA	2.3

Source: https://www.statista.com/statistics/270126/largest-stock-exchangeoperators-by-market-capitalization-of-listed-companies/

3.1. Stock Exchanges in India

3.1.1. Bombay Stock Exchange (BSE)

BSE is the world's fastest stock exchange with a median trade speed of 6 microseconds. It is the first Indian exchange to obtain certification from the World Federation of Exchanges in 2021. BSE also provides an online platform for trading in equities, derivatives, currency, and debt instruments.

3.1.2. National Stock Exchange of India (NSE)

NSE has introduced a variety of initiatives such as NSE Mobile App, NSE Now Trading Platform, NSE e-IPO platform and NSE Streaming Services. These initiatives have enabled investors to trade more conveniently and securely. It has also helped to reduce the cost of transaction and has made trading more accessible. The exchange has also been able to attract a larger number of investors and capitalize on the growth of the financial market. This has provided an opportunity for the exchange to further expand its customer base, allowing it to capture more market share. The exchange has also implemented strong security measures to protect investors from fraud and other malicious activities. Furthermore, the exchange provides comprehensive information and analysis, making it easier for investors to make informed decisions.

3.1.3. Multi Commodity Exchange of India (MCX)

Commodities traded on the MCX include bullion, metals, energy, and agricultural products. It is the largest commodity derivatives exchange in India. A commodities trading platform has been provided for the first time in India by this exchange. In addition, options trading in commodities has been launched for the first time on the exchange. More than 2,200 brokers and sub-brokers are authorized to provide trading services throughout the country. MCX has also launched a mobile trading platform for its users. It provides traders with real-time insights and trading opportunities. The platform also offers risk management tools and research reports to help traders make informed decisions [26].

3.1.4. National Commodity and Derivatives Exchange (NCDEX)

Securities and Exchange Board of India (SEBI) regulates NCDEX, one of the leading commodity exchanges in India. NCDEX is a platform for trading in commodities like pulses, cotton, guar seed, oil, oilseeds, and bullions. It ensures the safety of investments and promotes fair practices in the commodities market. It also provides market data and information for traders. Agriculture, metals, and industrial commodities are all available there. In addition to clearing, settlement, and counterparty risk management, NCDEX offers a variety of services to its members. Fair trading practices are also enforced by

NCDEX. Buyers and sellers can access the same price through a central limit order book. As a result of this system, prices are more stable and price discovery is larger [27].

4. Methodology

DWT is a signal processing technique used to decompose a signal into different frequency components. It is also used to reduce noise in signals by removing high-frequency components that are likely to contain noise. By using this technique for stock indices, it is possible to identify trends more accurately and identify potential market movements. The Haar wavelet function in range level 1 to level 5 was used in stock movements' analysis. The wavelet analysis was used to identify the short-term price movements and the model was developed and tested using the two indices. The results show that the model is able to accurately predict the short-term price movements of the two indices. The Sensex is composite formed by thirty companies stocks and Nifty 50 is a composite indicator of top fifty companies listed in national stock market. Apart from testing the two indices using judgemental sampling technique two companies each from the composite of Sensex and Nifty 50 were also subjected to forecasting to validate the model. For validating the model 1024 days stock price data were used for each of the index and selected stocks. This data were kept as training set and another five days data following the 1024 days were used for validation [28]. The training set data were denoised using Haar Wavelet at five levels. The correlation coefficient and MSE were taken as parameters to find out the optimum level at which the denoised signal matched the original level. The highest correlation coefficient and lowest mean square error indicated the optimum signal used for forecasting the data. For forecasting data linear model, general exponential model, polynomial models of degree one, two, three and general power of degree one and two were used. Coefficient of determination was used to decide the best model. The model with biggest coefficient was selected as appropriate model. Then the model was used to forecast the next day data. Finally Buy or Sell decision was recommended using formula.

5. Wavelet Denoising

Through the use of various window sizes, it is possible to obtain both time domain information as well as frequency domain information simultaneously in the wavelet transform through its property of multiresolution. An example of a wavelet-based denoising algorithm is shown in figure 3 and follows the basic idea behind it. There are three main steps involved in decomposing the model: first, modifying the detail coefficients, and second, reconstructing the model. De-noising is used when there is an assumption that the decomposition has been completed. Reconstruction methods can be directly performed if there is no need for any modification process. A reconstructed signal is made up of the original signal, the final level approximation, and the individual level details. It's called perfect reconstruction. Choosing the wavelet function to use in signal decomposition is the first step in wavelet de-noising. Wavelets can be divided into various types and subtypes. A candidate wavelet function is then used to select the appropriate

decomposition level. A threshold rule must be selected for the third step. A hard thresholding function and a soft thresholding function are usually used in thresholding algorithms. A wavelet processor's overall performance depends heavily on the choice of threshold. This work utilized HAAR and db2 wavelet functions for stock analysis as they are commonly used for de-noising purpose.

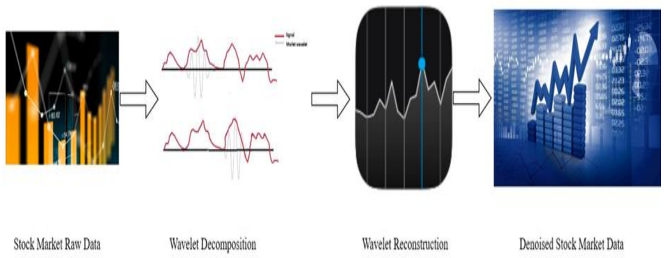

Fig. 1. Wavelet Denoising Procedure

At the end of the decomposition process, the low-frequency components of the signal are stored in the approximation coefficients, while the high-frequency components are stored in the detailed coefficients. This decomposition process is known as the Discrete Wavelet Transform (DWT).

5.1 NSE Nifty 50 Wavelet Analysis For De-Noising

NSE NIFTY 50 stock index data are taken for analysis in this section. The length of data considered for analysis is 2724 trading days' closing price. Since the input range to be considered for wavelet analysis should be in 2N (where N is the size of the array), out of the total data 256, 512 and 1024 trading days closing prices were taken for analysis for validating the best size using correlation of original and reconstructed data.

5.1.1 Wavelet Analysis of Trading Length 256 days of NSE NIFTY 50

The stock index NSE NIFTY 50 closing price during the period between 21.12 .2020 and 29.12.2021 for the purpose of keeping 28 trading days that is equal to 256 days were taken for analysis and the results were present in Table 2

Table 2. Analysis of 256 Trading Day's Data

Level	Threshold	Correlation	MSE
1	536.83	0.99523	7.4636
2	536.83	0.98704	3.1885
3	536.83	0.97915	1.0863
4	536.83	0.97386	0.0963
5	536.83	0.9708	0.4587

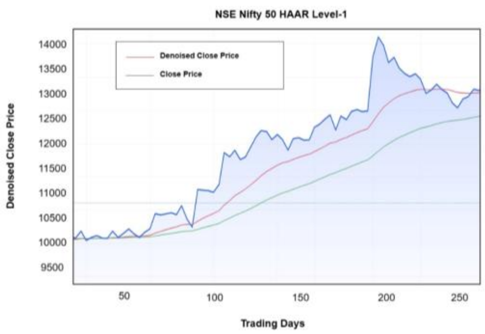

Fig. 2. Original and De-noised Signal at Level 1

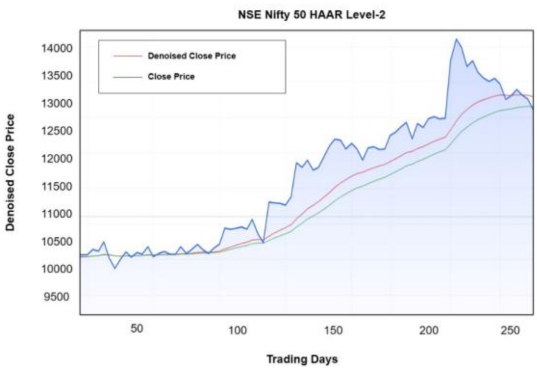

Fig. 3. Original and De-noised Signal at Level 2

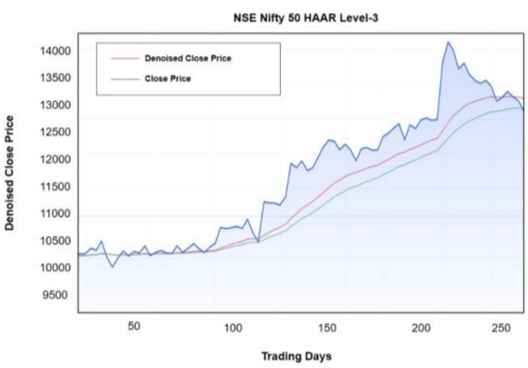

Fig. 4. Original and De-noised Signal at Level 3

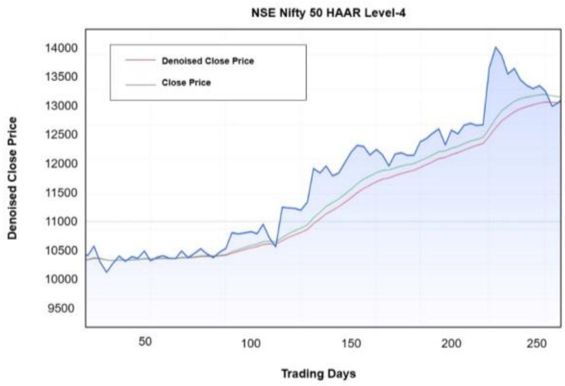

Fig. 5. Original and De-noised Signal at Level 4

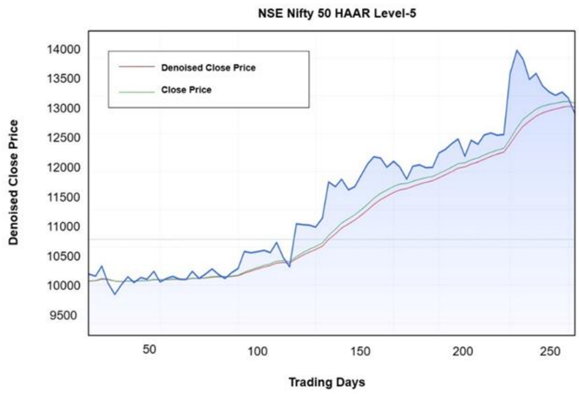

Fig. 6. Original and De-noised Signal at Level 5

The stock index NSE NIFTY 50 closing price of 256 trading days is taken for analysis. Based on the Figure 1 to 5 and table 2 it is inferred that the analysis resulted in a correlation measures of 0.99523, 0.98704, 0.97915, 0.97386 and 0.97078, and mean square of errors of 7.4636, 3.138, 1.0863, 0.9095 and 0.4795. Since the data pattern is very rough, the wavelet process is repeatedly applied. It is found that at level three the mean square error value is observed to be the least and hence, the for forecasting the fifth level denoised signals are selected. The main aim is to reduce the risk of over fitting in training phase. The comparison of the models used for evaluating the forecast of denoised signal, it is found that linear polynomial order of degree 3 best fit the data since the R2 value (0.9855) is the highest and Root mean square error (RMSE-78.05) is the lowest compared to all the other models.

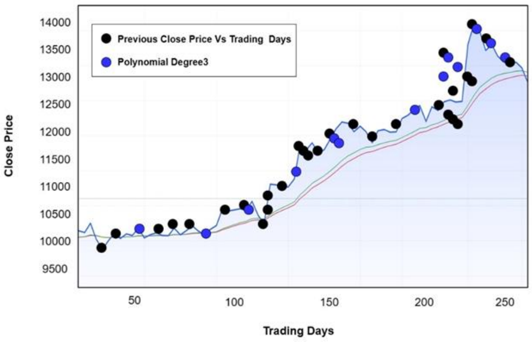

Fig. 7. Original and forecasted curves for the denoised data

Hence, for forecasting the model f(x) = 0.0001396*x^3 - 0.08134*x^2 + 21.03*x + 8037 is selected and forecasting is done for testing module which comprises of next five days data set. All data shown in Table 3.

Table 3. Forecast values and Actual values of Test Module

Sr. No	Time Value	Forecast	Actual
1	257th Day	10438.94	10.43
2	258th Day	10445.85	10.44
3	259th Day	10452.81	10.44
4	260th Day	10459.83	10.04
5	261st Day	10466.9	10.23

The methods selected for measuring the performance are root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and Theil's inequality coef cient (Theil U). The values of the four errors were calculated and tabulated in Table 4

Table 4. Analysis of Errors

Sr. No	Error	Error Value
1	MSE	6112.4
2	MAE	4112.1
3	MAPE	0.278
4	THEILU	0.116

5.2.2 Wavelet Analysis of Trading Length 512 days of NSE NIFTY 50

Table 5. Analysis of 512 Trading day's data

Level	Threshold	Correlation	MSE
1	536.83	0.99937	8.054
2	536.83	0.99845	4.099
3	536.83	0.99915	2.165
4	536.83	0.99386	1.322
5	536.83	0.99085	0.869

The stock index NSE NIFTY 50 closing price during the period between 08.12 .2020 and 29.12.2021 for the purpose of keeping 29 trading days that is equal to 512 days were taken for analysis and the results were present in Table 5.

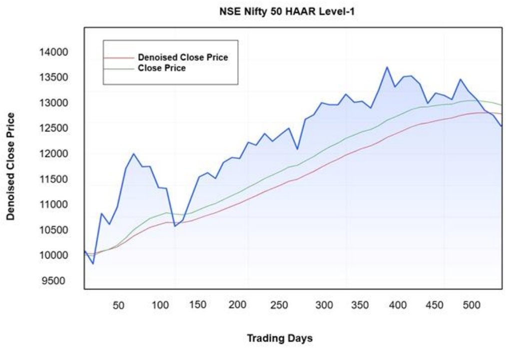

Fig. 8. Original and De-noised Signal at Level 1

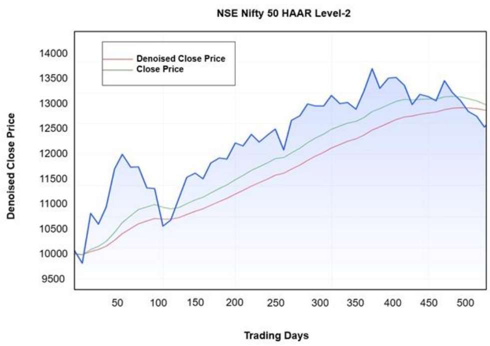

Fig. 9. Original and De-noised Signal at Level 2

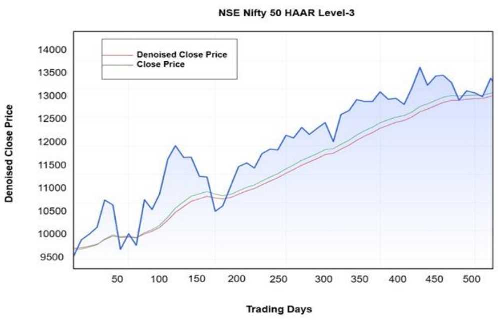

Fig. 10. Original and De-noised Signal at Level 3

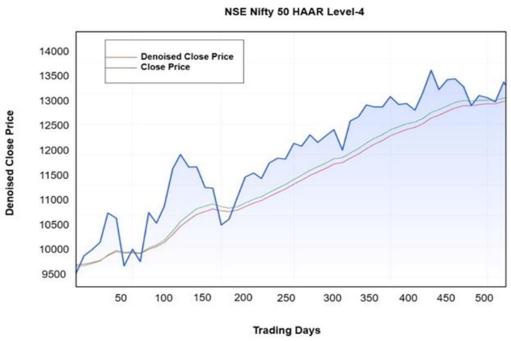

Fig. 11. Original and De-noised Signal at Level 4

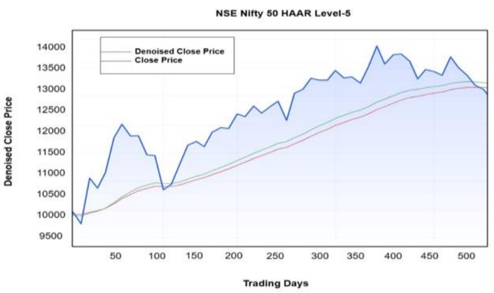

Fig. 12. Original and De-noised Signal at Level 5

From Table 5 and Figures 7 to 11 it is inferred that the stock index NSE NIFTY 50 closing price of 512 trading day's length is taken for analysis. The analysis resulted in correlation measures of 0.99937, 0.99845, 0.99758, 0.99706 and 0.99674 and mean square of errors of 8.0094, 4.0999, 2.1678, 1.3229 and 0.86493. The wavelet process is repeatedly applied because the data pattern is very rough. It is found that at level five the mean square error value is observed to be the least and hence, for the forecasting process the fifth level denoised signals are selected. The main aim is to reduce the risk of overfitting in training phase. The comparison of the models used for evaluating the forecast of denoised signal, it is found that linear polynomial order of degree 3 best fit the data since the R2 value (0.9191) is the highest and Root mean square error (RMSE-265.3) is the lowest compared to all the other models.

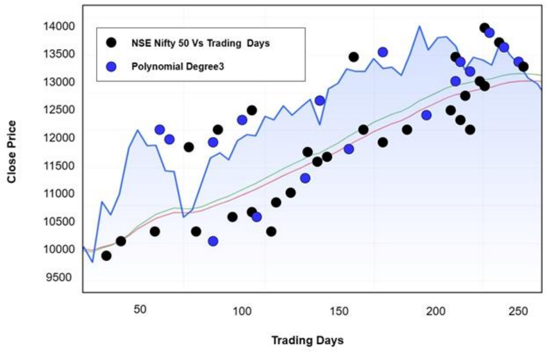

Fig. 13. Original and forecasted curves for the denoised data

Hence, for forecasting the model f(x) = 0.00000734 *x^3 - 0.001148*x^2 + 4.846*x + 7418 is selected and forecasting is done for testing module which comprises of next five days data set.

Table 6. Analysis of Errors for Forecasting for TL 512

Sr. No	Error	Error Value
1	MSE	5660.4823
2	MAE	5659.2323
3	MAPE	0.2003612
4	THEILU	0.091066

From Table 6 it is observed that the value of errors MSE, MAE, MAPE and Theil U are 5660.4823, 5,659.25, 0.200331 and 0.091066 respectively. These values were used as performance measures to find out the optimum length of trading days and to take decisions on trading by comparing it with other trading lengths.

6 Summary of Findings

The analysis of the stock index NSE NIFTY 50 closing price of 256 (21.12 .2016 and 29.12.2017) trading days resulted in correlation measures of 0.99523, 0.98704, 0.97915, 0.97386 and 0.97078, and mean square of errors of 7.4636, 3.138, 1.0863, 0.9095 and 0.4795. It is found that at level five the mean square error value is observed to be the least and hence, the for forecasting the fifth level denoised signals are selected. The fifth level reconstructed data was used for forecasting using various models. The comparison of the models used for evaluating the forecast of denoised signal, it is found that linear polynomial order of degree 3 best fit the data since the R2 value (0.9855) is the highest and Root mean square error (RMSE-78.05) is the lowest compared to all the other models. Hence for forecasting the model f(x) = 0.0001396*x^3 -0.08134*x^2 + 21.03*x + 8037 is used. The forecast values of the next five days were found to be 10438.94, 10445.85, 10452.81, 10459.83 and 10466.9. According to the calculations, its found that 16816.36, 16815.50, 9838, and 0.3297 in terms of mean square error, mean absolute error, and mean absolute percentage error. After analysing the major stock indices, two companies were selected from each of the index forming constituents to analyse the company's share price movements.

7 Conclusion and Future Directions

From the analysis and findings, it is concluded that Haar Wavelet used for denoising smoothening the price signals worked well for forecasting share prices. Further empirical result the analysis of selected indices and four companies confirmed the price movements with accuracy. The models, if used with care, can provide an accurate prediction of stock prices and help stock holders make informed decisions. This will help create a fair and competitive stock market, where stock holders can make the most of their investments and contribute to a stronger economy. The analysis consumed the interval period as one day but intraday variations were observed in share trading practices. This must be incorporated to make the analysis helpful for traders trading on margins. The next assumption is the discrete data behavior. This may be improved if large amount of data confirming normal behavior may be taken and analyzed continuous wavelet analysis. The static data behavior may be enhanced with dynamic analysis by incorporating deduction and induction of old and new values arrived from online data on a continuous basis. The Haar wavelet alone is incorporated in this work the other wavelet functions also be used to denoise the signals.

References

[1] Cheng, CH, Wei, LY & Chen, YS 2021, 'Fusion ANFIS models based on multi-stock volatility causality for TAIEX forecasting', Neuro Computing, pp. 3462-3468.

[2] Chen & Nai-Fu 2018, 'Financial Investment Opportunities and the Macroeconomy', Journal finance, vol. 46, no. 2, pp. 529-554.

[3] Christopher M Bilson, Timothy J Brailsford & Vincent J Hooper 2017, 'Selecting macroeconomic variables as explanatory factors of emerging stock market returns', Pacific Basin Finance Journal, vol. 9, no. 4, pp. 401-426.

[4] Nithya J & Tamizhchelvan, G 2016, 'Effectiveness of Technical Analysis in Banking Sector of Equity Market', IOSR Journal of Business and Management, vol. 16, issue. 7, pp. 20-28

[5] Ntim, CG, Opong, KK, Danbolt, J & Dewotor, F 2018, 'Testing the weak-form efficiency in African stock markets', Managerial Finance, vol. 37, no. 3, pp. 195-218. DOI: 10.1108/03074351111113289

[6] Osei 2019, 'Asset Pricing and Information Efficiency of the Ghanaian Stock Market', AERC Research Paper 115, Nairobi, Kenya.

[7] Parasuraman, K & Elshorbagy, A, 'Wavelet networks: an alternative to classical neural networks', Proceedings of IEEE International Joint Conference on Neural Networks, vol. 5, pp. 2674–2679.

[8] Pradeep Kumar, G & Saravanan, S 2019, 'Prediction of Stock Option Prices Using Volatility (Garch (1, 1)) Adjusted Black Scholes Option Pricing Model', Indian Journal of Finance, vol. 7, issue. 9 pp. 36-44

[9] Gencay, R, Selcuk, F & Whitcher, B 2021, 'An Introduction to Wavelets and Other Filtering Methods in Finance and Economics', Academic Press, New York.

[10] Sapankevych & Sankar, R 2020, 'Time Series Prediction Using Support Vector Machines A Survey', Computational Intelligence Magazine, vol. 4, no. 2, pp. 24-38.

[11] Shiguang Ma & Michelle L Barnes (2020), 'Are China's Stock Markets Really Weakform Efficient', Centre for International Economic Studies, Adelaide Australia Discussion Paper.119, pp.1-47.

[12] Soltani, S 2020, 'On the use of the wavelet decomposition for time series prediction', Neurocomputing, vol. 48, pp. 267-277.

[13] Tsung-Jung Hsieh, Hsiao-Fen Hsiao & Wei-Chang Yeh 2019, 'Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on arti cial bee colony algorithm', Applied Soft Computing, vol.11, pp. 2510–2525.

[14] Umaprabha, S& Malavika, M 2019, 'A Study on Technical Analysis of Stocks Listed in NSE with reference to Pharmaceutical Industries', International Journal of Management (IJM), vol. 6, pp. 79-86.

[15] Vinodkumar P Pathade 2019, 'Equity research: Fundamental analysis for long term investment', International Journal of Applied Research, vol. 3, issue 4, pp. 678-682.

[16] Abu-Mostafa, YS & Atiya, AF 2015, 'Introduction to financial forecasting', Applied Intelligence, pp. 205–213.

[17] Yao, J, Tan, C & Poh, HL 2017, 'Neural networks for technical analysis: A study on KLCI', International Journal of Theoretical and Applied Finance, vol. 2,

pp. 1-21.

[18] Refenes, AN, Zapranis, AD & Bentz, Y 2015, ‘Modelling Stock Returns With Neural Networks’, International Conference on 3D Digital Imaging and Modeling , vol.1, pp. 39-50.

[19] Ahmed S Wafia, Hassan Hassana & Adel Mabrouka 2015, ‘Fundamental Analysis Models in Financial Markets – Review Study’, Procedia Economics and Finance, vol. 30 pp. 939 – 947

[20] Amara Graps 2015, ‘An Introduction to wavelets’, IEEE Conference on Computational Science and Engineering, Summer, published by the IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720, USA., vol. 2, no. 2.

[21] Aronson & David, R 2016, ‘Evidence-Based Technical Analysis’, Hoboken, New Jersey: John Wiley and Sons, pp. 357, 355-356, 342. ISBN 978-0-470-00874-4.

[22] Berry, R & Sulochana 2017, ‘Study on Use of Technical Analysis in Forecasting Price Moments of Selected Companies of NSE & BSE’, Imperial Journal of Interdisciplinary Research, vol-3, Issue. 2, pp. 1394-1401.

[23] Blume, Lawrence, Easley, David & Maureen, O'Hara 2014, ‘Market Statistics and Technical Analysis’, The Role of Volume, Journal of Finance, vol. 49, issue. 1, pp. 153-81.

[24] Boopalan, C 2014, ‘Technical Analysis in Select Stocks of Indian Companies’, International Journal of Business and Administration Research Review, vol.2, issue.4, pp. 26-36.

[25] L. Anand et al., “Development of machine learning and medical enabled multimodal for segmentation and classification of brain tumor using MRI images,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–8, Aug. 2022. doi:10.1155/2022/7797094

[26] Deshpande Radhika, Bhalekar Dipali, Mutkule Prasad, Sanjay Pandhare, Nawale Akshay(2015) , “One Stop Solution for Farmer Consumer”, Interaction, IJCA Proceedings on National Conference on Advances in Computing NCAC.

[27] Mutkule Prasad R., “Interactive Clothing based on IoT using QR code and Mobile Application”, International Journal of Scientific Research in Network Security and Communication, vol. 6, issue-6, 2018.

[28] P. Mutkule and M. Ankoshe, “A survey on interactive clothing based on IOT using QR code and Mobile Application,” International Journal of Computer Sciences and Engineering, vol. 6, no. 6, pp. 652–654, Jun. 2018. doi:10.26438/ijcse/v6i6.652654