Introducing the Synergy Based Approach for Forecasting the Crude Oil Prices with Traditional and Machine Learning Econometric Models

Arslan Munir Turk, Saud Ahmed Khan, Muhammad Aamir and Zahanat Hussain ®

Submission Time: 21 May 2025 | **Acceptance Time:** 13 August 2025

ABSTRACT

Crude oil plays a pivotal role in the global economy, influencing inflation, trade balances, and energy security. Accurate forecasting of crude oil prices is therefore essential for policymakers and market participants. This study proposes a hybrid forecasting framework that synergizes conventional econometric methods with machine learning (ML) techniques. . First, the time series is decomposed using Ensemble Empirical Mode Decomposition (EEMD) to isolate intrinsic mode functions (IMFs). These components are then classified into deterministic and stochastic elements via spectral analysis. Second, traditional models such as ARIMA and GARCH are applied to the relevant IMFs, while advanced ML models (LSTM and XGBoost) are fitted to both original and residual series. Finally, a synergy model combines econometric and ML outputs, with Bayesian optimization applied for hyperparameter tuning. . Model performance is assessed using key error metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The findings suggest that hybrid models integrating conventional econometric methods with machine learning approaches, optimized through Bayesian techniques, achieve superior forecasting accuracy compared to standalone models. Additionally, the Diebold-Mariano (DM) test confirms that these synergy-based models offer the most reliable predictions for crude oil prices.

Key words: EEMD, Spectral Analysis, LSTM, XGBOOST, Hybrid Econometric Models

JEL Codes: C32, C45, C53, Q41, Q47

1. INTRODUCTION

Econometric analysis has the foundation of study and understanding of data-generating process to draw more accurate forecast and projections Stock and Watson (2015). For this, there is a variety of techniques, and based on these techniques, a series of models are tailored and being utilized for relevant fields and variables. In the modern era, Artificial Intelligence (AI) is being used as

PhD Student of Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad, (email: arslaneconometrician@gmail.com), Tel: +923336754342. ORCID: 0000-0002-3245-5232

® Professor of Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad, (email: saud@pide.org.pk), Tel: +923337028817, ORCID: 0000-0001-8458-6608

Professor of of Statistics, Abdul Wali Khan University, Mardan, (email: aamirkhan@awkum.edu.pk), Tel: +923339381636, ORCID: 0000-0003-1895-5350

Student of Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad, (email: zahanathussain1996@gmail.com), Tel: +923068015999, ORCID: 0000-0002-4733-7757

machine learning (ML) in almost every field of life. Researcher are using these ML models either neural network or non neural network based for forecasting the numbers Eşidir and Gur (2023). With the advancement of technology, the said techniques and models needed to be enhanced and equipped with ML. Integration of said conventional models and techniques with ML models and techniques not only facilitates and validates the existing as well as provides new insights that lead to more precise and sophisticated outcomes Mullainathan and Spiess (2017).

There is a process of selection of appropriate traditional models such as ARMA/ARIMA or ARFIMA or GRACH and machine learning like Kourentzes Neural Networks and ensemble methods (Hyndman and Athanasopoulos, 2018; Kourentzes et al., 2014). Before this, it is recommended to use the data exploratory analysis (EDA) for missing observations, outliers, and data transformation Panagiotelis et al. (2023). This helped to uncover the data patterns and characteristics of data sets. However, there are some issues in ML-based models and conventional econometrics models. Based on these issues, there would be fore-casting errors and misunderstood of data-generating processes Makridakis et al. (2018). There is always a need to decompose the time series data for analysis and forecasting. This involves the components which are trend, seasonality, cyclic, and irregularities. There are variety of techniques that are employed to de-compose it into its components. The classical decomposition is additive and multiplicative which have seasonal patterns Pana-giotelis et al. (2023). Moving average is another decomposition technique, it reduces the noise and highlights the trends by taking averaging values over a specified window Golyandina and Korobeynikov (2014). The seasonal trend decomposition uses LOESS (STL) to combine additive and multiplicative models and provide flexibility for seasonal and trend patterns as well as robustness to outliers Wu and Huang (2009). Wavelet is also the well-known method for decomposition, it uses wavelet transformation for the time series data decomposition into multiple frequency components. The said method is well effective for non-stationary data because it can capture the variations both in time and frequency.

EEMD stands for Ensemble Empirical Mode Decomposition which is an improved version of empirical Mode Decomposition (EMD) by incorporating white noise to fix the mode mixing issues. This decomposition divides the series into intrinsic mode functions (IMFs). It has significant importance for the analysis of non-stationary and nonlinear time series data through said IMFs which basically demonstrate the signals of time series data. It is very beneficial for econometric time series data due to this enhancement, where it provides precise signal decomposition for complex economic data and its modeling that leads to more precise prediction. Most recent research and studies demonstrated the efficacy of EEMD for capturing the trends and dynamic patterns of time series economic data. By this modeling, model is improved in terms of its robustness and reliability (Cai, et al., 2025; Guo et al. 2025). EEMD has also shown its significance as a tool for econometric modeling for interoperability and prediction power booted for integration with ML-based models such as machine learning and deep learning algorithms Niu et al. (2021). Further, spectrum analysis provided a better understanding of distinguish the deterministic and stochastic parts of decomposed series parts Umebayashi et al. (2017). For fitting and estimations, Autoregressive moving average (ARMA) models are more reliable and effective for linear data that has stationary characteristics. They are less reliable in the case of complex and non-linear data sets Box et al. (2015). Further, there is a need for fast and more flexible models and techniques to understand the complexity of data and non-linearity which can be fixed by ML models or techniques Hyndman and Athanasopoulos (2018). There are several ML-based models,

however, among the NN, the Long short-term memory (LSTM) is more popular for capturing the long-term effects and complexity of nonlinear data sets which makes them stronger than other time series applications Hochreiter and Schmidhu-ber (1997). Along with this attribute of LSTM, almost all ML-based models have issues with the data generating (black boxes) process, and their interoperability issue which makes them weaker than traditional models. This leads to less reliability of such models for underlying dynamics of the forecast Lipton (2018). To get better results from EEMD, tuning of LSTM hyper-parameters is required. For this, Bayesian optimization for the tuning of hyper-parameters has emerged as compared to traditional methods like random and grid research. This significant attribute helps improve efficiency and effectiveness for the navigation of hyper-parameter space. Specifically, Bayesian Optimization has the quality to build the probabilistic models for the objective function by exploring intelligently the search space while the random search evaluates random sample hyper-parameters and grid search exhaustively evaluates all combinations that can be possible. Bayesian optimization, allows the process to focus on the most promising areas, and reduction of evaluation to get the optimal hyper-parameters. As a result, Bayesian optimization can produce better performance with fewer computations with fewer resources, which makes it more advantageous for high-dimensional and complex hyper-parameters spaces of LSTM. Empirically, it has been demonstrated that Bayesian optimization not only accelerates the process of tuning but also enhances the predictive accuracy of the model through the identification of optimal configuration (Snoek et al., 2012; Bergstra et al., 2011). Along this, hyper-parameters selection for ML models is much caring for its performance, and for this batch size, number of units, and learning rate play significant roles. The rationale behind this limitation is that the traditional grid tuning for hyper-parameter is inefficient and its computation is expensive. To fix this issue, Bayesian provides a solution as an alternative through a probabilistic model of objective function, this may lead to more efficiency by minimizing the evaluation number Shahriari et al. (2015). LSTM may perform better through Bayesian optimization in terms of efficiency and systematic for its hyper-parameters tuning Frazier (2018). This is similar to a non-neural network based on another technique that is called XGBOOST. As discussed above, both types of models have their own limitations which are inherent in their nature. However, there is a need to combine the usage of such models for a better understanding of the data generating process and forecasting. Their combined usage may lead to the solution of the underlined issues. This combination would provide synergy among such types of models, address their limitations, and advance the forecast accuracy Zhang (2003). This approach enhanced the strength of ARIMA in terms of handling seasonality and trends, getting benefits from EEMD Assaad et al. (2008) and ML models, like LSTM and XG- BOOST. As a result, this synergy of said models provides a better understanding of the data and mitigates the forecasting errors.

2. RESEARCH OBJECTIVE

Conventional models (ARFIMA, GARCH) are less capable of dealing with nonlinearity issues often found in time series data, and consequent projections from these models might be misleading. Although machine learning tools such as LSTM, XGBOOST, etc. are efficient for nonlinear data but possess overfitting problems. Recently developed, Ensemble Empirical Mode Decomposition and Bayesian Optimization are expected to provide better projections on the basis of the less restricted structure and theoretical assumptions of these tools.

The objective of this research is to devise a synergy of conventional econometric and ML tools

through the EEMD technique for better forecasts/projections which will be more useful after the Bayesian algorithm for hyper-parameter optimization of underline machine learning models. Further, IMFs analysed separately to identify the determinsitc and stochastic parts as an objective to better fitting the ARIMA models.

3 REVIEW OF LITERATURE

Time series data are worth, and forecasting for different disciplines like economics, finance, climate science, and engineering are being used for production and policy making. For forecasting, there are two approaches used in the current era which are conventional and Machine Learning based. The primary conventional model is the Auto Regressive Integrated Moving Average (ARIMA) and in Machine Learning, these are Long Short-Term Memory (LSTM) and XGBOOST. Research are using both type of models as standalone for find the best model through comparison of their results of for forescasting Eşidir et al. (n.d), Eşidir et al. (2022). ARIMA was introduced by Harvey and Todd (2012) and is widely accepted for the usage of time series data generating process model identification. It has the capacity to capture seasonal patterns and linear trends for stationary time series data Box et al. (2015). Moreover, ARIMA has limitations like not being suitable for nonlinear and non-stationary data sets, less capability to handle multivariate data, and sensitivity to parameter selection (Hyndman and Athanasopoulos, 2018; Özer and Wei, 2006). LSTM was introduced by Hochreiter and Schmidhuber (1997) and is acceptable all around the world due to computing non-linearity relationships and is more popular for long-run dependencies in time series data. LSTM has peak strengths in language processing, time series forecasting, and speech recognition as well as (Hochreiter and Schmidhuber, 1997; Siami-Namini et al. 2018). There are some issues on the LSTM side, which are computational complex system (black box), generalized interpretation, and parameters tuning (Greff et al. 2020; Lipton, 2018). There are some studies those have recently explored the synergy (integration) of ARIMA and LSTM, these discussed the robustness of this synergy. Among those, Zhang, (2003) discussed the hybrid model of ARIMA with NN in terms of the combination of linear and nonlinear models for time series data. For instance, Assaad et al. (2008) data generating process through ARIMA for linear components and for nonlinear LSTM being used to improve the forecasting (Assaad et al. 2008 and Siami- Namini et al. 2018). Interestingly, the LMD-based ARIMA and LSTM proposed by Nasir et al. (2023) for better forecasting of crude oil prices. However, hyperparameter tuning can be improved through further Bayesian optimization. This research proceed same optimization process for XGBOOST as well as Zhao et al. (2023). In conclusion, With the limitations and advantages of both types of models, their synergy (integration) can present a new and better way with Bayesian optimization as well as EEMD for data generating process and forecasting. The spectral analysis is a reliable tool for examining behavior of each IMF through spectral results to consider as deterministic or stochastic part of a decomposed series Roushangar and Alizadeh (2018). Bayesian is the widely used powerful method for the optimization of hyper-parameter tuning of machine learning methods such as LSTM and XGBOOST specifically for the time series data (Firoozabadi et al. 2024; Olaniyan et al. 2024).

For fine-tuning (optimization) of the hyper-parameter's of XGBOOST max-depth, learning rate, and the number of estimators (estimators), the Bayesian optimization method used that is the surrogate model and has an iterative procedure for improvement performance. The studies have shown for the better performance and accuracy of this method than other methods like grid or

random search, especially for financial and energy data forecasting Xiong et al. (2022). Bayesian optimization application for LSTM networks is well known for its ability for time series data. The hyper-parameters are defined as the number of layers, hidden units, size of batch, and dropout rates. This method of optimization helps identify optimal configurations, enhancing prediction accuracy as well as reducing over-fitting. This researcher has successfully applied time series forecasting for energy and financial markets Malakouti et al. (2024).

4. RESEARCH METHODOLOGY

The new approach used for this research is a synergy of Conventional and ML-based models. The proposed research refers to the approach which is introduced by Nasir et al. (2023) as the synergy underlines both types of models. After collecting the data, the conventional methods for model identification of data generating process observed, for time series box-Jenkins methodology followed.

There are many ways for data decomposition processes like TCSI, LMD, EEMD, etc. I adopted EEMD as it is a new approach and less time-consuming. This research followed Nasir et al. (2023) for data generating and forecasting strategy which can be found in detailed in their research paper with the replacement of EEMD instead of LMD along with Bayesian optimization for hyper-parameter tuning. After this, spectral analysis was conducted to identify the stochastic or deterministic decomposed components (IMFs) of the series. This identification is based on energy and frequency of the IMF, more energy and less frequency is likely to be deterministic. The threshold is defined as if IMFs has more than 50% energy and less than 10% frequency then IMF will be deterministic, else will be stochastic (Shafiq et al. 2019 and Fernández and Sicre, 2008). Following this, each component is separately examined for to fit an ARIMA for estimation as well as adding stochastic and deterministic components are added and these combined models are fitted. However, after the model is fitted, their residuals integrate with the ML model's LSTM and XGBOOST. These ML models also fit directly to the original series and examined their results using accuracy metrics (RMSE, MSE, MAPE) and train test validation by splitting the data (cross-validation) with a 70 and 30 percent ratio. After that, it examined how the synergy is useful for DGP and forecasting. For hyper-parameter optimization, Bayesian Optimization (MSE is an objective function) is used for both LSMT/XGBOOST econometric models. As first step, series decomposed with EEMD and later apply sprectral analysis to observe for stoachastic or deterministic identofocation on each IMF. After that ARIMA model is fitted on deterministic and stochastic IMFs. After ARIMA models application, residuals will be integrated with machine learning models which are XGBOOST and LSTM for synergy of conventional and machine learning models. After integration, the model estimated as stack XGBOOST and LSTM , another stpe applied lated through Bayesian algorithm for their parameters estimation. ARIMA, LSTM, and XGBOOST also be fitted standalone for comparison. All models are compared with loss fucntions, their forecasting accuracy is compared through the Diebold Mariano test for the accuracy of the model. During the whole process, some extream values were observed which were leading to extream mean absolute percentage aerrors which were adressed using MAPE as symmetric mean absolute percentage error Makridakis et al. (2018).

Following are the mathematics and algorithms that have been used in this proposed research.

4.1. Ensemble Empirical Mode Decomposition

This research used the advance version of empirical mode decomposition that is Ensemble Empirical Model Decomposition (Huang and Wu, 2007; Wu and Huang, 2009 and Torres et al. 2011).

Decomposition Statistics:

$$y_t = \sum^{N} IMF_i(t)$$

where y(t) is the original time series data $IMF_i(t)$ are Intrinsic Mode Functions obtained from the decomposition.

4.2. Spectral Analysis

Energy computation for each IMF

$$E_i = \sum^{T} IMF_i^2(t)$$

Frequency computation through Fourier series

$$Y(k) = \sum^{T} y_t e^{-i2\pi ft} \qquad \text{k=0,1,2,----,T-1}$$

The above two components used for spectral analysis based on energy and frequency of each IMF.

4.3. Conventional Model Arima

$$\phi(B)(1-B)^d y_t = \mu + \theta(B) \in_t$$

4.4. Machine Learning Model Lstm

LSTM: (forget gate$(f_t) = \sigma(W_f x_t + U_f h_{t-1} + b_f)$, input gate$(i_t) = \sigma(W_i x_t + U_i h_{t-1} + b_i)$, $C_t = tanh(W_c x_t + U_c h_{t-1} + b_c$, cell state gate$(C_t) = f_t \odot C_{t-1} + i_t \odot C_t)$, output gate$(O_t) = \sigma(W_o x_t + U_o h_{t-1} + b_o)$, hidden state update $(h_t) = O_t \odot \tanh(C_t))$

4.5. Machine Learning Model Xgboost

XGBOOST: $\hat{y_i} = \sum_{i=1}^{K} f_k(x_i)$,
Objective function: $\mathcal{L}=\sum_{i=1}^{n} \iota(y_i, \hat{y}_i)+\Omega(f_k)$

4.6. Hyperparameters Tuning Through Bayesian Optimization

For synergy of ARIMA and LSTM EEMD-SD-ARIMA-LSTM-Bayesian: Optimization of hyper parameters through Bayesian algorithm: $x^*=argmin_{x \in X} f(x)$

Gaussian Process (GP) model: $f(x) \sim GP(\mu(x), \delta(x))$
X={$x_1 \in (10,200)$, $x_2 \in (0.0001,0.1)$, $x_3 \in (0.0,0.5)$, $x_4 \in (16,128)$}
x_1=number of LSTM units, x_2=learning rate, x_3=dropout rate, x_4=batch size

For synergy of ARIMA and XGBOOST

EEMD-SD-ARIMA-XGBOOST-Bayesian:optimization of hyper parameters through Bayesian algorithm:

$x^* = argmin_{x \in X} f(x)$

Gaussian Process (GP) model:

$f(x) \sim GP(\mu(x), \delta(x))$

X= {$x_1 \in (0.01,0.03)$, $x_2 \in (3,15)$, $x_3 \in (50,100)$}
x_1=estimated learning rate, x_2=maximum depth tree, x_3=number of boosting rounds

4.7. Model Accuracy Testing

The models compared through loss functions MAE, RMSE and MAPE. Moreover, model forecasting accuracy tested through Diebold Mariano test (DM) for significance difference of models with less loss functions. The statistics of DM test is widely used for accuracy testing especially for hybrid and forecasting model comparison;

$$DM = \frac{\bar{d}}{\sqrt{[\gamma_o + 2\sum_{k=1}^{h-1}\gamma_k]/n}}$$

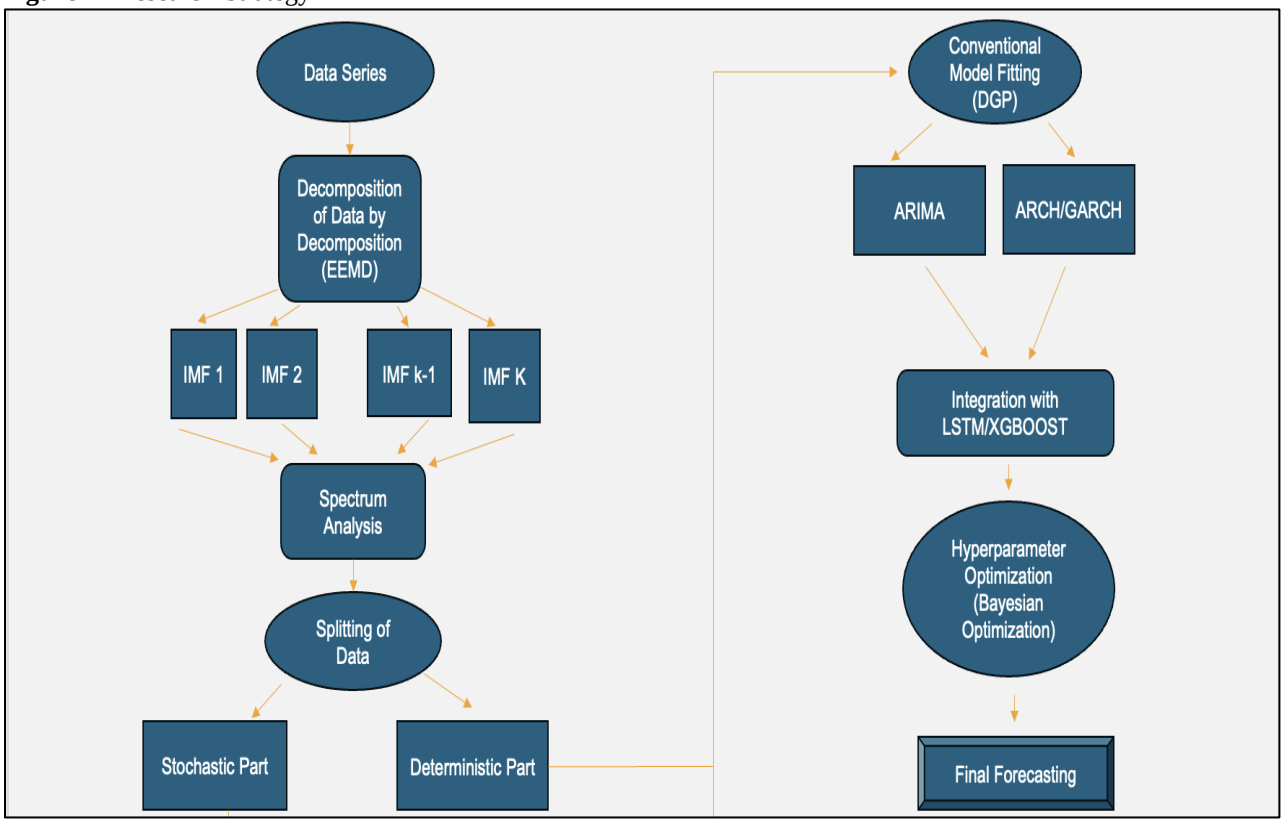
Where:

$\bar{d}$= Mean loss differential
n= Sample size
γ_k= Autocovariance at lag k
h= Forecast horizon
$\sum_{k=1}^{h-1}\gamma_k$=long run variance adjustment
γ_o=Variance of the loss differential

5. RESEARCH STRATEGY

The intention is to introduce this research design as it is a combination of conventional econometric models and ML-based models that yield robust results both in terms of data generating process and forecasting. Further, it is the synergy of conventional and ML-based techniques to mitigate the underlying issues and enhances the capacity of data for better forecasting. All models are compared with each other based on their losses and evaluatedsimple, synergy based and optimized models for best performance.The following figure depicts the proposed design.

Figure 1 Research Strategy



6. DATA

Data was gathered from World Trade Index website from January 01, 2018 to December 31, 2024. The frequency of this data is daily for crude oil prices.

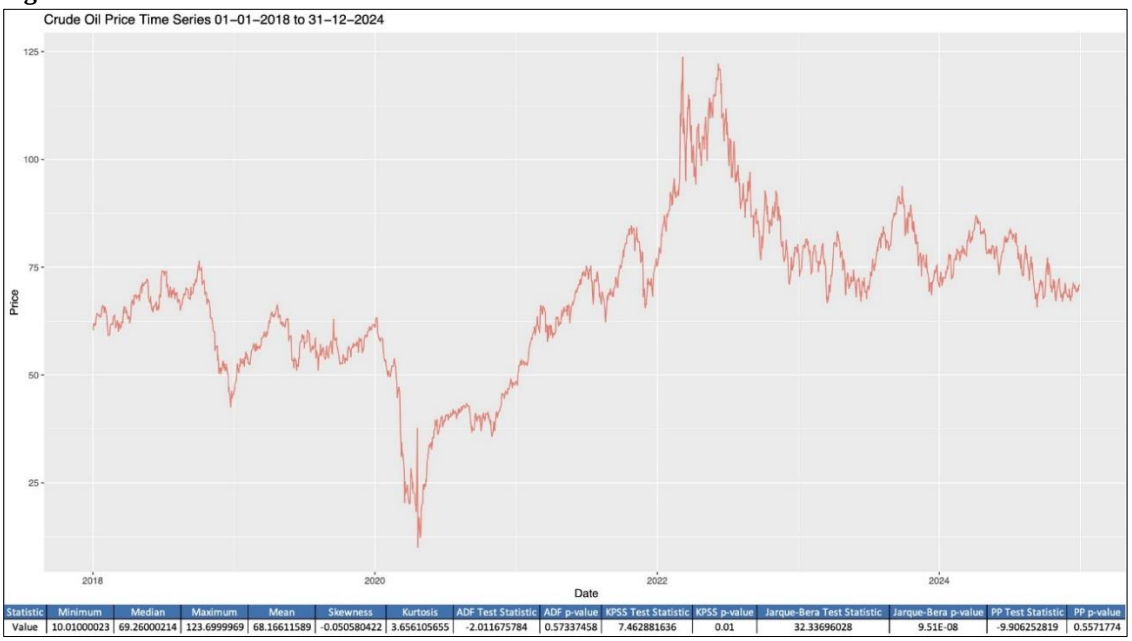
7. ANALYSIS

The analysis is based on each standalone each method/model which is discussed above and synergy achieved through the proposed technique. Based on the model's fitting and their forecasting checks through MAE, RMSE, and MAPE. After this process, this research may be able to tell if this synergy is more reliable or standalone strategy is better by using the Diebold Mariano test for model forecasting accuracy.

8. RESULTS AND DISCUSSION

The following results are presented in order of methodology and proposed models. Below Figure 2 presents basic statistics of underline data. The data is showing a fluctuated attitude of prices of crude oil and there is a major shift during covid era (2020-2021) as well

Figure 2 Series Basic Info

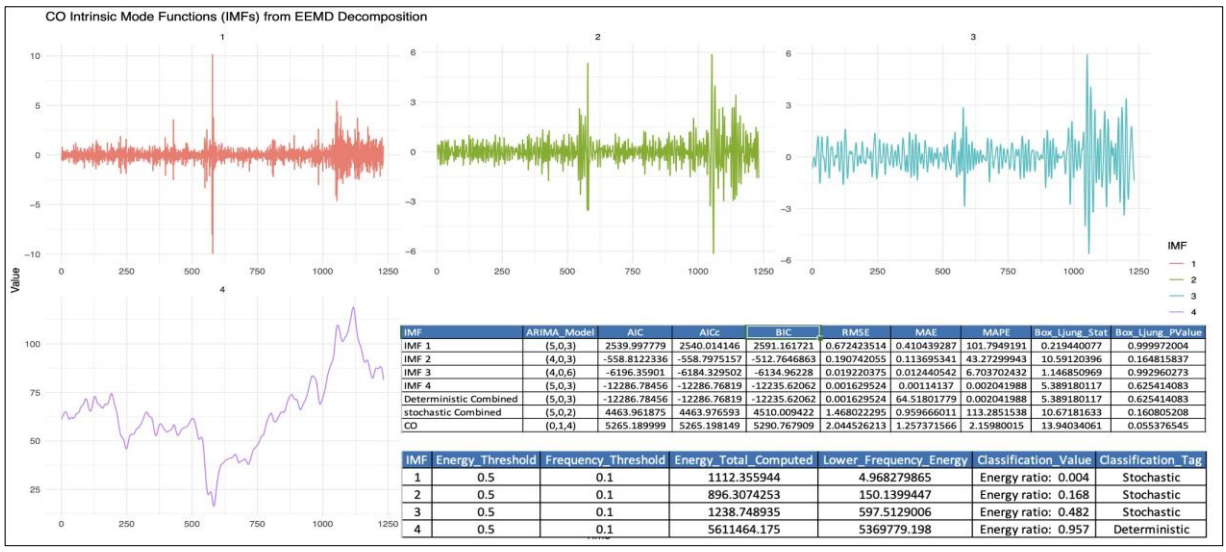


Statistic	Minimum	Median	Maximum	Mean	Skewness	Kurtosis	ADF Test Statistic	ADF p-value	KPSS Test Statistic	KPSS p-value	Jarque-Bera Test Statistic	Jarque-Bera p-value	PP Test Statistic	PP p-value
Value	10.01000023	69.26000214	123.6999969	68.16611589	-0.050580422	3.656105655	-2.011675784	0.57337458	7.462881636	0.01	32.33696028	9.51E-08	-9.906252819	0.5571774

Figure 3 demonstrates the decomposition-based graphical results of intrinsic mode functions (IMFs) through the ensemble empirical mode function method. There are total of 4 IMFs inclusive of the residual. Further, information on the original series, combined stochastic and combined deterministic series information also presented as well. Their ADF test statistics along with p-value also presented right below the Figure 3.

As per estimation methodology, the IMFs further analyzed based on spectral analysis for more clarification to determine the deterministic and stochastic IMFs as shown below in Figure 3.Figure 3 EEMD Decomposition of Series.

Figure 3 EEMD Decomposition of Series



IMF	ARIMA_Model	AIC	AICc	BIC	RMSE	MAE	MAPE	Box_Ljung_Stat	Box_Ljung_PValue
IMF 1	(5,0,3)	2539.997779	2540.014146	2591.161721	0.672423514	0.410439287	101.7949191	0.219440077	0.999972004
IMF 2	(4,0,3)	-558.8122336	-558.7975157	-512.7646863	0.190742055	0.113695341	43.27299943	10.59120396	0.164815837
IMF 3	(4,0,6)	-6196.35901	-6184.329502	-6134.96228	0.019220375	0.012440542	6.703702432	1.146850969	0.992960273
IMF 4	(5,0,3)	-12286.78456	-12286.76819	-12235.62062	0.001629524	0.00114137	0.002041988	5.389180117	0.625414083
Deterministic Combined	(5,0,3)	-12286.78456	-12286.76819	-12235.62062	0.001629524	64.51801779	0.002041988	5.389180117	0.625414083
stochastic Combined	(5,0,2)	4463.961875	4463.976593	4510.009422	1.468022295	0.959666011	113.2851538	10.67181633	0.160805208
CO	(0,1,4)	5265.189999	5265.198149	5290.767909	2.044526213	1.257371566	2.15980015	13.94034061	0.055376545

IMF	Energy_Threshold	Frequency_Threshold	Energy_Total_Computed	Lower_Frequency_Energy	Classification_Value	Classification_Tag
1	0.5	0.1	1112.355944	4.968279865	Energy ratio: 0.004	Stochastic
2	0.5	0.1	896.3074253	150.1399447	Energy ratio: 0.168	Stochastic
3	0.5	0.1	1238.748935	597.5129006	Energy ratio: 0.482	Stochastic
4	0.5	0.1	5611464.175	5369779.198	Energy ratio: 0.957	Deterministic

Figure 4 Accuracy results – MAE

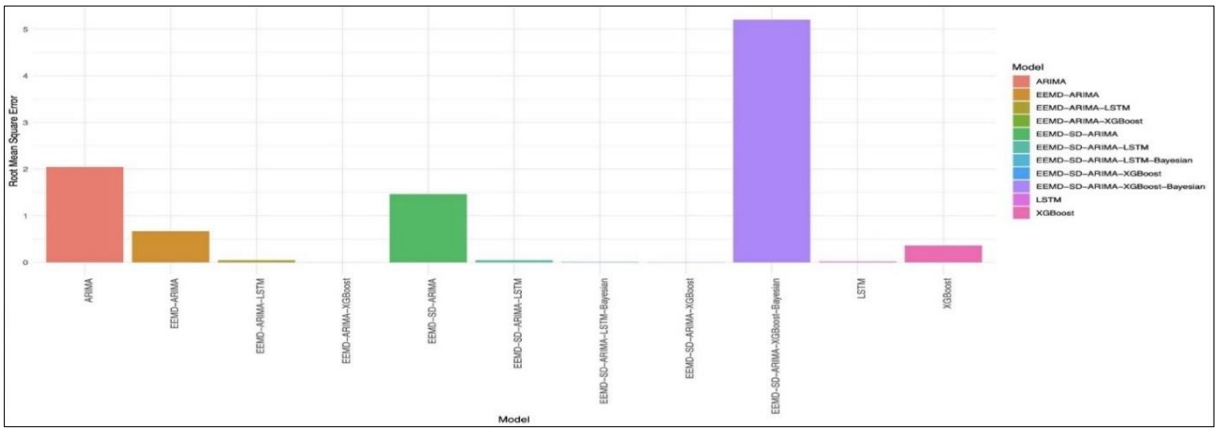


Figure 5: Accuracy results – RMSE

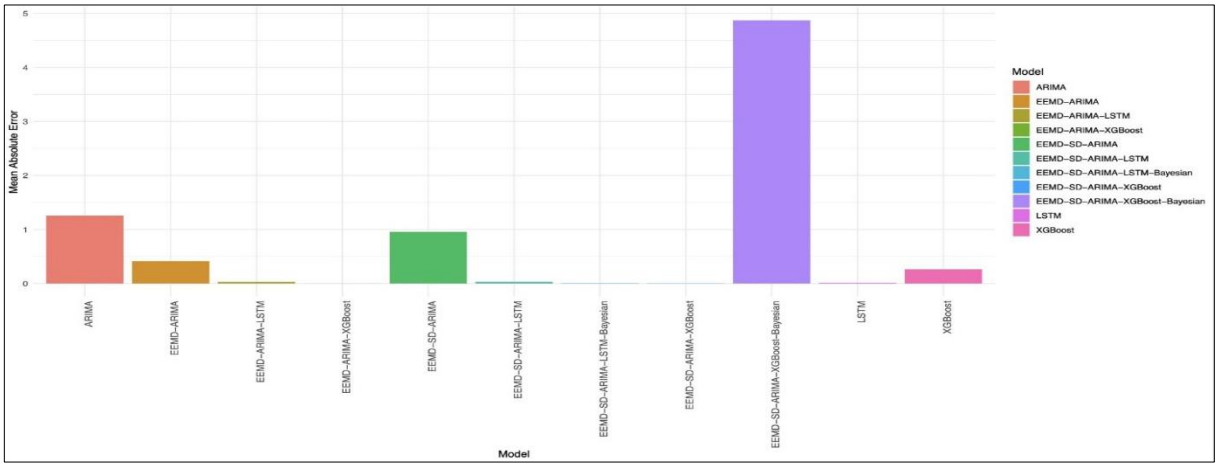
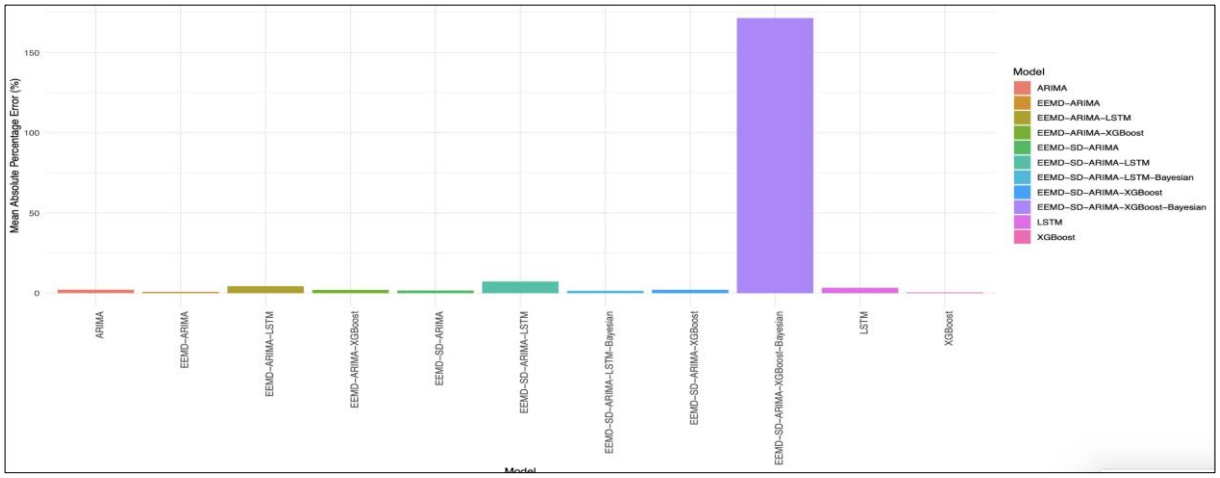


Figure 6: Accuracy results – MAPE



Based on spectral analysis for further clarification was obtained to determine the deterministic and stochastic IMFs as shown below in Figure 3. With the help of this, models were built for estimation and comparison. Figure 4, Figure 5, and Figure 6 depict the errors (MAE, RMSE, MAPE) as a bar chart for a better visual representation of the magnitude of errors. By observing all models, the models based on the synergy of conventional econometric and machine learning models have more worth than others as their errors are less than any other model. Other models are standalone as conventional econometric models or machine learning models. Further, the Bayesian optimization is also applied before forecasting through these models for their hyperparamete.

9. SIGNIFICANCE OF RESEARCH

The expected finding of this study is to explore explored the significance of the synergy of Conventional and ML-based econometrics. The EEMD and Bayesian hyperparameter Optimization criterion paves the new way forward towards forecasting accuracy. Further, the proposed synergy between both types of forecasting techniques (ML and Conventional Models) provided better estimates for comparison of standalone and synergy results. Conventional and ML-based synergy provide robust data-generating processes and more accurate forecasting, The next step is to determine whether the results of the proposed hybrid and synergy are stable, consistent, better, and with lower MAE than existing regression models. Further, the testing of this synergy technique applied for different data sets (economic and financial data sets) may produce better forecasting and have a detailed analysis of black box exploration of ML-based econometrics model as well. This helps to see the detailed issues among conventional and ML-based methods as standalone as well as synergy among them to produce precise and more accurate results. The technique may help researchers and policymakers to better understand time series data and further planning for the future. As per the proposed techniques for better data-generating understanding and better forecasting, the following are the best final models with the synergy of conventional and ML-based econometric models:

- EEMD-SD-ARIMA-LSTM (Bayesian Optimized)
- EEMD-SD-ARIMA-XGBOOST (Bayesian Optimized)

where:

- EEMD: Ensemble Empirical Mode Decomposition
- SD: the combination of stochastic deterministic components after decomposition
- ARIMA: Auto Regressive Integrated Moving Average
- LSTM: Long Short-Term Memory
- XGBOOST: Extreme Gradient Boosting

10. CONCLUSION

From the above errors and accuracy results of different models, it has been observed that synergy-based models have lower errors than standalone models. The Table 1 below conatains the numerical values from the above three figures for further details while Table 2 presnets the model accuracy test statistics for the underlying crude oil price series based on the proposed research. The p-value of the DM test for synergy-based and Bayesian-optimized model shows its worth as

the best model among all other models as listed.Overall, XGBOOST and LSTM synergy with ARIMA models showing more worth than other models with less errors (MAE, RMSE, MAPE). The synergy-based model EEMD-SD-ARIMA-LSTM-Bayesian has least errors than any other model. As shown that some of the models were unable to address the criteria of minimum MAE, RMSE and MAPE , EEMD-SD-ARIMA-XGBoost-Bayesian model is one of them which has significantly high value of MAPE which showing high sensitivity with extreme values. It has consistently higher erros for MAE, RMSE and MAPE as compare to other models either standalone or synergy based optimized mdoels. Similarly ,the EEMD-ARIMA-XGBOOST has comaptively low errors than green highlighted which is EEMD-SD-ARIMA-LSTM-Bayesian but the formar model has greater MAPE than the later model.

Model	MAE	RMSE	MAPE
ARIMA	1.257371566	2.044526213	2.15980015
LSTM	0.012190058	0.018436539	3.385610058
XGBoost	0.261227069	0.359722444	0.452559702
EEMD-ARIMA	0.414960604	0.667793834	0.802535038
EEMD-SD-ARIMA	0.956103749	1.462368365	1.678944483
EEMD-ARIMA-XGBoost	0.001503112	0.001996276	2.051181327
EEMD-SD-ARIMA-XGBoost	0.003987271	0.005337506	2.103189338
EEMD-ARIMA-LSTM	0.029928821	0.047681514	4.335823212
EEMD-SD-ARIMA-LSTM	0.030958968	0.047397056	7.28649608
EEMD-SD-ARIMA-XG Boost-Bayesian	4.872038142	5.197231694	171.4786216
EEMD-SD-ARIMA-LSTM-Bayesian	**0.006466643**	**0.009969837**	**1.511782808**

Table 1: Accuracy Results

Model Comparison	Test Stat	P Value
ARIMA vs LSTM	20666.97333	0
ARIMA vs XGBoost	7.535939178	9.39E-14
LSTM vs XGBoost	-19.03367643	5.67E-71
ARIMA vs EEMD-XG Boost-ARIMA-SD-Bayesian	5785.893155	0
LSTM vs EEMD-XGBoost-ARIMA-SD-Bayesian	-21.85510884	1.02E-89
ARIMA vs EEMD-XGBoost-ARIMA-SD	19777.88622	0
ARIMA vs EEMD-ARIMA-XGBoost	18616.44986	0
EEMD-ARIMA-XGBoost-SD-Bayesian vs EEMD-ARIMA-SD-XGBoost	19.84967527	2.72E-76
EEMD-ARIMA-SD-LSTM-Bayesian vs EEMD-ARIMA-SD-LSTM	-0.081254312	0.935252937
EEMD-ARIMA-SD-LSTM-Bayesian vs EEMD-ARIMA-SD-XGBOOST	-4.227173659	2.54E-05

Table 2: DM Test Results for Model Accuracy

11. LIMITATIONS

This study is focused on synergy-based results, not on individual techniques, the proposed synergistic approach was found more superior to the others compared in this study. However, the study, synergy was used for univariate techniques not the multiple regression techniques or ML models for multiple variables forecasting. This is major limitation and remains a gape to prove the approach through multiple regression and ML techniques.

REFERENCES

Assaad, M. , R. Boné and H. Cardot (2008). "A New Boosting Algorithm for Improved Time-Series Forecasting with Recurrent Neural Networks." *Information Fusion* 9(1):41-55.

Bergstra, J. S., R. Bardenet, Y. Bengio and B. Kégl (2011). "Algorithms for Hyper-Parameter Optimization." *Advances in Neural Information Processing Systems* 24.

Box, G. E, G. M. Jenkins, G. C. Reinsel and G. M. Ljung (2015) "Time Series Analysis: Forecasting and Control 5th Edition." *Journal of Time Series Analysis,* 712.

Cai, X. , D. Li , J. Zhang and Z. Wu (2025) "MA-EMD: Aligned Empirical Decomposition for Multivariate Time-series Forecasting." *Expert Systems with Applications* 267.

Eşidir , K. A. , Y. E. Gur and V. Yoğunlu (n.d). "Estimating Monthly New Car Sales in Turkey with Artificial Neural Networks (ANN) and ARIMA Models." *Pamukkale University Journal of Business Research* 9(2):260-277.

Eşidir, K. A. and Y. E. Gur. (2023) "Turkish Plastic Industry Import Estimate With Artificial Neural Networks." *Akademik Hassasiyetler* 10(23):91-114.

Eşidir, K. A. , Y. E. Gür , V. Yogunlu and M. Çubuk (2022) "Estimating Monthly New Car Sales in Turkey with Artificial Neural Networks (ANN) and ARIMA Models." *Üniversitesi İşletme Araştırmaları Dergisi* 9(2):260-277.

Fernández, F G and J. E. Sicre (2008) "On the Use of Empirical Mode Decomposition for Detecting Deterministic Dynamics." *Physica D: Nonlinear Phenomena* 237(5):685-701.

Firoozabadi, S. S. , M. Ansari and F. Vasheghanifarahani (2024). "Crude Oil Trend Prediction During COVID-19: Machine Learning with Randomized Search and Bayesian Optimization." *European Journal of Business and Management Research* 9(3):6-13.

Frazier, P. I. (2018). "A Tutorial on Bayesian Optimization." *arXiv preprint arXiv* 930.

Golyandina, N. and A. Korobeynikov (2014). "Basic Singular Spectrum Analysis and forecasting with R." *Computational Statistics & Data Analysis* 71:934-954.

Greff, K. , S. van Steenkiste and J. Schmidh. (2012). "On the Binding Problem in Artificial Neural Networks." *arXiv preprint arXiv* 208.

Guo , Y. , J. Si, Y. Wang, F. Hanif, S. Li, M. Wu, M. Xu and J. Mi (2025). "Ensemble-Empirical-Mode-Decomposition (EEMD) on SWH Prediction: The Effect of Decomposed IMFs, Continuous Prediction Duration, and Data-driven Models." *Ocean Engineering* 324.

Harvey, A. C. and P. H. J. Todd (2012). "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study." *Journal of Business & Economic Statistics* 299-307.

Hochreiter, S. and J. Schmidhuber (1997). "Long Short-Term Memory." *Neural computation* 9(8):1735-1780.

Huang, N. E and Z. Wu (2007). "A Review on Hilbert-Huang Transform: Method and Its Applications To Geophysical Studies." *Reviews of Geophysics* 1-23.

Hyndman, R.J. and A.B. Koehler (2006). "Another Look at Measures of Forecast Accuracy." *International Journal of Forecasting* 22(4):679–688.

Hyndman, R.J., and G. Athanasopoulos (2018). *Forecasting: Principles and Practice*. OTEXTS.

Kourentzes, N. , D. K. Barrow and S. F. Crone (2014). "Neural Network Ensemble Operators for Time Series Forecasting." *Expert Systems with Applications* 4235-4244.

Lipton, Z. C. (2018). "The Mythos of Model Interpretability: In Machine Learning, The Concept of Interpretability is Both Important and Slippery." *Queue* 31-57.

Makridakis, S. , E. Spiliotis and V. Assimakopoulos (2018). "Statistical and Machine Learning Forecasting methods: Concerns and Ways Forward." *PloS one* 13(03).

Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018). "Statistical and Machine Learning forecasting methods: Concerns and ways forward." *PLOS ONE* 13(3):e0194889.

Malakouti, S. M., F. Karimi, H. Abdollahi, M. B. Menhaj, A. A. Suratgar and M. H. Moradi (2024). "Advanced Techniques for Wind Energy Production Forecasting: Leveraging Multi-layer Perceptron+ Bayesian Optimization, Ensemble Learning, and CNN-LSTM Models." *Case Studies in Chemical and Environmental Engineering* 10:100881.

Mullainathan, S. and J. Spiess (2017). "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31:87-107.

Nasir, J. , M. Aamir, Z. Ul Haq, S. Khan, M. Y. Amin and M. Naeem (2023). "A New Approach for Forecasting Crude Oil Prices based on Stochastic and Deterministic Influences of LMD Using ARIMA and LSTM Models." *IEEE* 14322-14339.

Niu, Z. , G. Zhong and H. Yu (2021). "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing* 48-62.

Olaniyan, J., D. Olaniyan, I. C. Obagbuwa , B. M. Esiefarienrhe, A. A. Adebiyi and O. P. Bernard (2024). "Intelligent Financial Forecasting with Granger Causality and Correlation Analysis Using Bayesian Optimization and Long Short-Term Memory." *Electronics* 13(22):4408.

Özer, Ö. and W. Wei (2006). "Strategic Commitments for an Optimal Capacity Decision under Asymmetric Forecast Information." *Management Science* 1238-1257.

Panagiotelis, A., P. Gamakumara, G. Athanasopoulos and R. J. Hyndman (2023). "Probabilistic Forecast Reconciliation: Properties, Evaluation and Score Optimisation." *European Journal of Operational Research* 306(02):693-704.

Roushangar, K. and F. Alizadeh (2018). "Entropy-based Analysis and Regionalization of Annual Precipitation Variation in Iran During 1960–2010 Using Ensemble Empirical Mode Decomposition." *Journal of Hydroinformatics* 20 (2):468-485.

Shafiq, M. S., T A Cheema, A A Khawaja and R k. (2019). "IMF Selection Using Energy and Dominant Frequency Features for Improved EEMD-based Signal Classification," *IEEE Access*." 7: 45185-45197.

Shahriari, B., K. Swersky, Z. Wang, R. P Adams and N. de Freitas (n.d). "Taking the Human Out of the Loop: A Review of Bayesian Optimization." *Proceedings of the IEEE* 104(1):148-175.

Siami-Namini, S. , N. Tavakoli and A. S. Nami (2018). "A Comparison of ARIMA and LSTM in Forecasting Time Series." *17th IEEE International Conference on Machine learning and Applications (ICMLA).* IEEE. 1394-1401.

Snoek, J., H. Larochelle and R. P Adams (2012). "Practical Bayesian Optimization of Machine Learning Algorithms." *Advances in Neural Information Processing Systems* 25.

Stock, J. H and M.W. Watson (2015). *Introduction to Econometrics (3rd Updated Edition).* https://swh.princeton.edu/~mwatson/Stock-Watson_3u/Students/RTC/Stock_Watson_3U_AnswersToReviewTheConcepts.pdf.

Torres, M.E., M.A. Colominas, G. Schlotthauer and P. Flandrin (2011). "A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4144-4147.

Umebayashi,K., M. Kobayashi and M.López-Benitez (2017) "Efficient Time Domain Deterministic-Stochastic Model of Spectrum Usage." *IEEE Transactions on Wireless Communications* 17(3):1518-1527.

Wu , Z. and N. E Huang (2009). "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method." *Advances in Adaptive Data Analysis* 01:1-41.

Xiong, X. , X. Guo, P. Zeng, R. Zou and X. Wang (2022). "A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization." *Frontiers in Energy Research* 155.

Zhang, G Peter (2003). "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model." *Neurocomputing* 50:159-175.

Zhao, H. , Y. Wang, X. Li, P. Guo and H. Lin (2023). "Prediction of Maximum Tunnel Uplift Caused by Overlying Excavation Using XGBoost Algorithm with Bayesian Optimization." *Applied Sciences* 13 (17): 9726.