Scope
Volume 16 Number 01 March 2026

Hyperparameter Tuning of Machine Learning Models for Time-Series Forecasting Using Metaheuristic Optimisation Algorithms

Wendy Ling Shin Yie[1], Siti Nur Iqmal Binti Ibrahim[1,2],
Nur Haizum Abd Rahman[3], Yong Ge[1]

Corresponding Author:**Yong Ge**

Abstract: Time-series forecasting is an essential component of data analysis and predictive modelling. The primary objective of time-series forecasting techniques is to minimise the percentage prediction error associated with future observations as much as possible. To achieve this goal, numerous statistical approaches and machine learning models have been proposed to capture temporal patterns and generate reliable forecasts.However,traditional statistical forecasting models often suffer from significant limitations in capturing long-time dependencies, addressing abrupt fluctuations, and dealing with non-linear interactions.To overcome these shortcomings, machine learning approaches have emerged as a powerful tool that has fundamentally revolutionised data analysis in recent years.Nevertheless, their forecasting performance remains sensitive to hyperparameter configuration, whereas manual tuning across high-dimensional hyperparameters can often be time-consuming, inefficient and suboptimal. Consequently, automatic hyperparameter optimisation (HPO) methods have been increasingly adopted. Among these, metaheuristic optimisation algorithms have emerged as innovative and powerful optimization techniquesand are currently among the most widely employed methods, with primary attention given to hyperparameters associated with machine learning architectures. By optimising the hyperparameters of machine learning models, these algorithms can substantially enhance forecasting accuracy and generalisability. This study reviews the performance of different metaheuristic optimisation algorithms and machine learning forecasting approaches, and further proposes a classification of metaheuristic optimisation algorithms and provides a clearer direction for the future design and implementation of hyperparameter optimisation strategies for deep learning models in time-series forecasting.

Keywords: statistical forecasting models; machine learning; metaheuristic optimization algorithms; hyperparameter optimisation.

Introduction

Time-series forecasting has long been recognised as a highly valuable field of research, and its wide applications have accordingly attracted considerable attention across many domains. With the time series approach to forecasting, historical observations of the same variable are analysed in order to construct a model that describes the underlying relationship (Siami-Namini et al., 2018). The established model is then employed to extrapolate the time series into the future. This modeling approach is particularly useful when little is known about the underlying data generating process or when there is no satisfactory explanatory model that relates the forecasting variable to other explanatory variables. Over the past several decades, much effort has been devoted to the development and improvement of time series forecasting models as well as the optimisation algorithms. Although a wide variety of time series forecasting methods have been proposed in the existing literature, they can be classified into two overarching categories: traditional statistical approaches and machine-learning-based methods.

These traditional statistical models face notable limitations when applied to time-series forecasting because they struggle to effectively capture the non-linear relationships (Pierre et al., 2023; Zhu et al., 2022) and long-term dependencies present in financial data (Bai et al., 2023), and cope with extreme volatility, resulting in suboptimal forecasting performance. To address these challenges, machine learning methods have emerged as powerful alternatives. Among them, deep learning models, characterised by multiple layers and complex structures, are better equipped to extract the underlying information embedded in time series. Nevertheless, the forecasting performance of deep learning models is sensitive to hyperparameter configuration, and optimal performance can only be achieved through careful hyperparameter tuning. Hyperparameters are predefined deep learning model parameters that determine both the network architecture before training and the training process prior to model fitting. However, manual hyperparameter tuning can often be labor-intensive and inefficient, particularly given the high dimensionality of the parameter space. Consequently, hyperparameter optimisation (HPO) has become a crucial solution for automating the search for optimal hyperparameters. This process generally involves defining an objective function, specifying the hyperparameter search space, and analysing various characteristics in order to determine the most suitable hyperparameter optimisation technique. Owing to their flexibility and gradient-free mechanisms, metaheuristic optimisation algorithms have become important tools for solving complex hyperparameter tuning problems (Abd Elaziz et al., 2021). By providing a systematic review of traditional statistical models, deep learning models, and various metaheuristic optimisation algorithms, this review offers practical directions for the future design and implementation of metaheuristic optimisation algorithms in the field of time-series analysis.

Traditional Statistical Approaches

Within the traditional forecasting approaches, one of the most widely used method of time-series forecasting is the family of linear stochastic models described by Box and Jenkins (1970), including AutoRegressive (AR), Moving Average (MA), AutoRegressive-Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA), and Seasonal AutoRegressive Integrated moving average (SARIMA) models. These approaches are now commonly referred to as either the Box-Jenkins methodology, or ARIMA models. As a culmination of the research by many prominent statisticians, the Box-Jenkins method has remained a standard time series forecasting approach over recent decades (Siami-Namini et al., 2018) and has been extensively applied to forecasting problems across a wide range of disciplines, including the social sciences (Benvenuto et al., 2020), economics (Mao et al., 2022), meteorology (Dabral & Murry, 2017), traffic flow (Kumar & Vanajakshi, 2015), stock markets (Dadhich et al., 2021), and so on.

To demonstrate that this family of models is not only theoretically well-grounded and interpretable but also empirically effective across forecasting tasks in diverse domains, a substantial body of evidence has been reported in the literature. For example, Valipour et al. (2013) utilised ARMA and ARIMA models to predict the inflow to the Dez dam reservoir and increased the number of parameters in order to improve predictive accuracy to four parameters. The results clearly showed that ARIMA$(4,1,1)(1,0,1)_{12}$ model achieved a lower RMSE of 0.7148 than ARMA$(0,4)(3,0)_{12}$ model, which yielded an RMSE of 0.7981. Ho and Xie (1998) proposed applying the ARIMA model to repairable system failure analysis and forecasting, and compared its performance with that of the Duane model. The results were satisfactory, with the mean absolute deviation (MAD) of the ARIMA model being substantially smaller than that of the Duane model, at 4.1 and 53.4 respectively. Nyangarika et al. (2018) proposed the modified ARIMA model in which exponential smoothing was used to support parameter estimation and forecasting. The results of the study stated that the forecasting model achieved an R^2 value of 0.97, suggesting a strong goodness of fit and explanatory ability of the model.

For seasonal time series data, short run non-seasonal components and significant autocorrelation at seasonal lags are also likely to contribute to the model. In such cases, it is appropriate to estimate a seasonal ARIMA model that incorporates both non seasonal and seasonal factors within a single specification. Dabral and Murry (2017) utilised the SARIMA model to forecast regional monthly, weekly and daily monsoon rainfall and obtained satisfactory results. The p-values of the Ljung-Box statistics clearly indicated that no further modelling was required, suggesting that the SARIMA model had captured the main features and information of the data very well. Ray et al. (2021)

employed a SARIMA model to analyse and forecast the monthly average rainfall and temperature of South Asian countries. The diagnostic checking of all selected models has found white noise by the Ljung-Box Q test and normally distributed residuals, suggesting that these SARIMA model provided a good fitness to the data. The forecasting results further showed that rainfall exhibited a declining trend, whereas temperature displayed a positive and increasing trend for all countries throughout the year. Kumar and Vanajakshi (2015) employed SARIMA model, historical average method, and naive method to forecast vehicular traffic flow. The results clearly showed that the corresponding MAPE values were 10.53 and 10.42 for the historical average method and naive method, respectively, both of which were higher than the MAPE of 9.22 obtained with the proposed SARIMA model.

Even though ARIMA models are widely used for modelling economic and financial time series, these models have some major limitations. For instance, it is hard for a simple ARIMA model to capture the nonlinear relationships between variables (Siami-Namini et al., 2018). Furthermore, ARIMA model typically assumes a constant standard deviation in errors, which in practice is often violated. When an ARIMA model is integrated with a GARCH model, this assumption can be relaxed. Mohammadi and Su (2010) used four alternative classes of GARCH models to characterise the behaviour of weekly oil returns and conditional variance. These findings broadly suggest that Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH) forecasts outperform those from the GARCH, Exponential Generalised Autoregressive Conditional Heteroskedasticity (EGARCH), and Fractionally Integrated Generalised Autoregressive Conditional Heteroskedasticity (FIGARCH) models, and the MA(1)-EGARCH(1,1) and MA(1)-APARCH(1,1) models are preferred. Pandey et al. (2019) applied the Box-Cox transformation to stabilise variability of rainfall time-series variance, thereby improving the forecasting performance of time-series model, and then employed a SARIMA-GARCH model to eliminate the conditional variance of SARIMA model of rainfall time series in two different climatic environments. These results additionally indicate a superior fit of the SARIMA-GARCH models compared to SARIMA model.

These previous studies suggest that no single method consistently dominates across all time series forecasting tasks, because the forecasting performance of model depends on the characteristics of the sample. Grubb and Mason (2001) clearly illustrate this point clearly by forecasting air passenger traffic at relatively long lead-times, on the order of ten years. They found that the forecasting performance of the ordinary Holt-Winters exponential smoothing model is generally less well than ARIMA or basic structural models. However, dampening the trend towards an average of some past values can improve the accuracy of prediction. They noted that ARIMA forecasts tended to be less smooth and accurate than those produced by Holt-Winters exponential smoothing model after incorporating a long-term average trend into the predictions. In financial analysis, non-stationarity, linear and non-linear dynamic correlation, volatility clustering, long-term dependencies, and structural breaks are significant features of financial time series. Under the pre-assumption that the conditional mean follows a linear dynamic structure (Zhang, 2003) and the conditional variance is updated through a parametric recursive equation (Bollerslev, 1986), the ARIMA-GARCH model has long been a dominant traditional approach to modelling financial returns and volatility. However, as financial markets have become more complex and the available information set has expanded, the limitations of this model in dealing with strong non-linearity, long-term dependencies, time varying parameters, and multi-source features have become increasingly evident (Pierre et al., 2023; Zhu et al., 2022), prompting subsequent research to introduce deep learning models such as Deep Neural Network (DNN) models. With multiple non-linear hidden layers, DNN models are superior at modelling complex relationships and are therefore well suited to addressing these forecasting challenges (Bharti et al., 2023). Aggarwal et al. (2022) utilised ARIMA, GARCH and DNN models to landslide forecasting. The results clearly showed that the DNN achieved the lowest RMSE, at 1.03×10^{-1}, whereas the ARIMA model obtained a substantially larger error of 3.29×10^{3}. Also, the MSE is higher for ARIMA model compared with DNN. Lou et al. (2022) employed the ARIMA model, the DNN model and the multivariate Long Short-Term Memory (LSTM) model to predict the disease burden of pneumoconiosis in Tianjin. The comparison of forecasting results indicated that both the multivariate LSTM model and the DNN model substantially outperformed the traditional ARIMA model.Thence,the existing literature generally utilises traditional statistical models as benchmark methods to evaluate theperformance of newly proposed approaches.

Machine Learning

The advent of machine learning approaches has fundamentally revolutionised financial market analysis. A DNN model is a type of machine learning method that draws inspiration from the deep structures of human perception and production systems. With continual advances in neural network models and training optimisation algorithms, the capacity of DNN model to process diverse data types, including images (Bai et al., 2014), text (Gandhi et al., 2023), and audio (Dwijayanti et al., 2018), has been enhanced. As neural networks are widely regarded as universal function approximators, it is natural to apply them to modelling seasonality and non-linearities in trend components. In principle, the Universal Approximation Theorem implies that, under certain conditions, neural networks with a certain structure can approximate any continuous function, complex patterns and relationships in data to any desired degree of accuracy (Csáji, 2001; Cybenko, 1989; Hornik et al., 1989). This implies that any lack of success in applications must arise from inadequate learning, insufficient numbers of hidden units or the lack of a deterministic relationship between input and target. Gorr (1994) also articulated that the neural network model form, with its non-linear threshold functions has the ability to concurrently identify almost any kind of seasonality and non-linear trends within the data. Moreover, Franses and Draisma (1997) also further emphasised that the hidden layer units in neural network models can convey information about potential fluctuations change in seasonal patterns. Thus, as one of the subsets of neural networks, DNN should possess a similar property to capture the seasonal variations and non-linear mappings in time-series data. Hasib et al. (2024) utilised a DNN model for electric vehicle driving range prediction, optimised using the RMSPropoptimiser. This proposed model achieved an R^2 of 0.99 and a MAPE of 2.01%, substantially outperforming conventional machine learning techniques such as support vector machines and linear regression.

DNN encompasses various types, approximately including Artificial Neural Network (ANN) with multiple hidden layers, Convolution Neural Network (CNN), Graph Neural Network (GNN), and Recurrent Neural Network (RNN). ANN originated from efforts to emulate the human brain. Schmidhuber (2015) denoted that ANN forms the conceptual foundation of DNN, comprising interconnected artificial neurons that collaboratively process and learn from data. Parthiban et al. (2007) were among the first to utilise ANN to estimate the cycle life characteristics of Li-ion cell with CoO anodes, achieving the best fit values corresponding to an error factor of <1% and a correlation coefficient of 0.98. CNN model consists of convolution layers and pooling layers that performed hierarchical feature extraction, which is contributed to information filtering, limited translation, robustness and rotation invariance, allowing it adept at reliably recognising features even when the input data are shifted, rotated or scaled (Mo & Zhao, 2024; Pawar & Jadhav, 2018). GNN is designed to operate on graph-structured data by iteratively aggregating and propagating information along edges, enabling the model to capture complex relational dependencies in non-Euclidean domains (Wu et al., 2021). Liu et al. (2017) claimed that the inputs to DNN can be presented in any order without any effect on the network performance. However, such architectures neglect the time-series nature and long-term dependencies in the temporal data. Meanwhile, the original ANN model also exhibits notable limitations. In particular, it lacks an explicit memory or state mechanism and suffers from time-consuming training, slow convergence, and a tendency to become trapped in local optima (Yuan et al., 2020). While these variants are effective in their respective domains, this study focuses on time-series forecasting, where preserving temporal nature and capturing non-linear relationships and long-term dependencies are central requirements. To exploit the temporal information embedded in time-series data more effectively, RNNs have been developed on the basis of ANNs and have demonstrated strong predictive performance. While RNN, as defined by Coşkun et al. (2017), is a neural network architecture primarily used for patterns detection in sequential data. It can capture information in previous computations and exploit these to further improve the accuracy of its outputs. Cao et al. (2012) utilise a multivariate RNN model to predict wind speed. The findings indicated that the multivariate RNN model significantly outperformed both the univariate RNN model and ARIMA model, with a reported MAPE of 11.5. Liu et al. (2020) proposed a real time regional collision risk prediction framework that combines Density-Based Spatial Clustering of Applications with Noise (DBSCAN) technique, Shapley value method and RNN model to generate forecasts at different time points within the selected water area. The results clearly suggest that the proposed approach can effectively predict collision risk in the study region, with prediction bias reported to be negligible. RNN model and its variants are generally more effective and accurate than the original DNN (El Makkaoui et al., 2024; Zraibi et al., 2022), CNN and GNN for modelling sequential information (Usharani, 2023). In view of this, RNN and its variants are more appropriate for analysing time-series data.

Although the RNN model has the notable advantage of its memory capability for modelling sequential patterns, it is persistently plagued by the vanishing and exploding gradient problems. Specifically, RNN no longer continues effective learning and fails to retain long-range information when dependencies span many time steps. One of common challenges is the long-term dependency problem, which arises when the gradients are used to update the network's weights during RNN training can become extremely small (vanishing) or excessively large (exploding) (Bengio et al., 1994; Coşkun et al., 2017). These gradient problems hinder the network's ability to effectively learn and capture long-term dependencies in sequential data so that the long past learning results are wiped from certain levels or cannot be utilised. Therefore, to address these limitations, LSTM model, an advanced variant of RNN, was developed by Hochreiter and Schmidhuber (1997). The LSTM model introduces specific gating mechanisms to mitigate the vanishing and exploding gradient problems, even under substantial time lags, resulting in more stable optimisation and better performance (Kang et al., 2024; Sahu et al., 2023; Siami-Namini et al., 2019; Zhu et al., 2022). Thus, LSTM-based architectures have been widely adopted for time-series prediction tasks in numerous fields, such as financial systems (Wu et al., 2018), energy systems (Pierre et al., 2023), meteorological systems (Haider et al., 2019), agricultural systems (Ray et al., 2023), traffic flow analysis (Hu et al., 2022), landslide research (Duan et al., 2023), medical research (Yang, Eb., et al., 2022; Zhu et al., 2022), etc. To assess the benefits of gating mechanisms, several studies have compared LSTM model with RNN model in time-series forecasting tasks. By way of illustration, Shahani et al. (2022) empirically compared LSTM, simple RNN, and RFA for forecasting the Drilling Rate Index (DRI). The results clearly illustrate that the LSTM revealed the best predictive performance for DRI with the highest R^2, the lowest RMSE, the largest VAF and an appropriate a-20 index values as 0.998, 0.19479, 0.996, and 0.997 in the testing stage, respectively, which substantially outperforms the simple RNN model.

Beyond comparisons with simple RNN model, it is crucial to assess whether LSTM model provides consistent improvements over traditional statistical forecasting approaches. Sirisha et al. (2022) applied ARIMA, SARIMA, and LSTM to sales forecasting and reported accuracies of 93.84% (ARIMA), 94.378% (SARIMA) and 97.01% (LSTM) approximately. These Results clearly suggest that LSTM model surpasses the traditional statistical models in constructing the best model. Siami-Namini et al. (2018) compared the forecasting performance of LSTM model and ARIMA model in terms of the minimisation of prediction error rates. These findings indicate that LSTM model consistently outperformed ARIMA model, achieving an average error rates reduction of approximately 84% to 87%, thereby highlighting superior predictive capability for historical monthly financial and economic time series across different periods. Duan et al. (2023) presented ARIMA model, a univariate LSTM model, and a multivariable LSTM model to predict the displacement of the Baijiabao landslide. This study illustrates that the multivariate LSTM model produced more accurate displacement forecasts than both ARIMA model and the univariate LSTM model, achieving the lowest error metrics (MSE = 0.64223 mm, RMSE = 0.801 mm, and MAE = 0.505 mm). Accordingly, these results clearly show that the multivariable LSTM model is proposed has higher accuracy and better application prospects. Banerjee et al. (2022) compared a multivariate LSTM with linear regression, KNN, regression tree, RF, SVR, ARIMA, and Markov switching model for long-term vegetable price forecasting. The results clearly indicate that the ANN model performs the worst, whereas the multivariate LSTM model achieves the greatest predictive accuracy, reporting RMSE of 1.13, MAE of 1.00, and MAPE of 1.71. These empiricalstudies have consistently reported superior forecasting accuracy for LSTM in complex time-series forecasting (Bharti et al., 2023; Haider et al., 2019; Liang et al., 2022; Zhu et al., 2022).

Furthermore, some studies attempt to combine LSTM with other methods to further enhance predictive performance. For example, Pierre et al. (2023) utilised hourly data from the Electric Company of Benin (CEB), collected over four years (2017-2021), to deduce peak-hour electricity consumption. This research compared the predictive performance of the traditional statistical models, deep learning methods, and hybrid approaches. The results indicate that the ARIMA model significantly underperformed relative to the deep learning and hybrid models, with the values of RMSE and MAPE being 49.90 and 12.05%, respectively. By contrast, the LSTM model achieves an RMSE of 18.74 and a MAPE of 3.01%, which substantially outperforms traditional models. Moreover, the hybrid ARIMA-LSTM model demonstrated the greatest accuracy, yielding an RMSE of 7.35 and a MAPE of 1.52%, followed by the hybrid ARIMA-GRU model, with the values of RMSE and MAPE being 9.60 and 1.56%, respectively. Yang, Eb., et al. (2022) integrated a multivariate multi-step LSTM model with the SHapley Additive exPlanations (SHAP) method to develop a new interpretable system for forecasting tuberculosis incidence, providing both local and global explanations of the multivariate single-step LSTM model incidence prediction. These findings demonstrate that, relative to the ARIMA benchmark, the multivariate two-step LSTM model achieved short-term reductions of 12.92% in RMSE, 15.94% in MAE, 15.97% in MAPE and 14.81% in sMAPE. Furthermore, the 3-step ARIMA-LSTM hybrid model delivered more excellent performance, reducing the corresponding errors by 15.19%, 33.14%, 36.79%, and 29.76%, respectively. Ray et al. (2023) introduced an improved hybrid ARIMA-LSTM model incorporating the random forest lag selection criterion, whereby a random forest is used to select LSTM input lags, thereby improving the robustness and predictive capability of the proposed model. These results clearly show that the proposed model proved its superiority over the conventional statistical models and LSTM model, yielding reductions of approximately 8-25% in RMSE, 2-28% in MAPE and 2-29% in MASE. Kang et al. (2024) incorporated the L2 regularisation term into the loss function when training LSTM model to prevent overfitting phenomenon. The results reveal that the L2-LSTM algorithm substantially outperforms RNN, L2-RNN, and conventional LSTM in terms of RMSE, MAPE, and the standard deviation of errors in both the training and testing. Vidal and Kristjanpoller (2020) utilised the hybrid CNN-LSTM model to forecast the gold volatility. These results indicate that a substantial improvement relative to both GARCH and LSTM models, with the proposed hybrid model achieving a 37% reduction in MSE is observed compared to the classic GARCH model and 18% compared to the LSTM model.

However, science continuously advances, and mathematicians have proposed exploiting the future observations for backward learning to further improve model accuracy (Kang et al., 2024). This has led to the development of the Bidirectional LSTM (Bi-LSTM) model, which adds an additional layer of LSTM that processes data from the future to the past, capturing the underlying relationships in time-series data. Bi-LSTM model always outperforms the predictive accuracy of unidirectional LSTM model, particularly in predicting the middle portions of time-series sequences (Kim & Moon, 2019) and capturing the long-term dependencies (Ubal et al., 2023). In practical applications, this bidirectional mode can help model more accurately capture patterns and anomalies within time-series data. For instance, Siami-Namini et al. (2019) utilised the Nikkei 225 Index, NASDAQ Composite Index, Hang Seng Index, S&P 500 Commodity Price Index, Dow Jones Industrial Average and IBM stock datasets to evaluate the predictive performance of ARIMA, LSTM, and Bi-LSTM models. These findings notably demonstrated that the Bi-LSTM model achieved an average RMSE of 20.17, compared with 39.09 for the unidirectional LSTM model and 302.96 for the ARIMA model. Overall, these results highlight the considerable potential of Bi-LSTM model for further development and innovation in financial forecasting and suggest that, for time-series analysis, Bi-LSTM model may be preferable to LSTM model for forecasting problems. Mardjo and Choksuchat (2024) also proposed a hybrid Bi-LSTM model for Bitcoin price forecasting. The results clearly show that, when combined with the ARIMAX GARCHX model, the Bi-LSTM variant significantly outperforms other LSTM variants in this time-series forecasting task. Amini and Kalantari (2024) proposed a hybrid CNN-Bi-LSTM model, incorporating automatic parameter tuning, to predict the daily closing gold price. These results indicate that the proposed hybrid approach significantly outperforms other models, including stacked-LSTM, CNN, CNN-LSTM, and ConvLSTM, in terms of total bias, capturing extreme values and obtaining promising results with RMSE=34.87, RMAE=5.15, R^2=0.95.

In summary, the existing studies suggest that LSTM and Bi-LSTM models offer substantial advantages for modelling time-series data.

Metaheuristic Algorithms

Despite their effectiveness, the performance of LSTM and BiLSTM models is highly dependent on the selection of hyperparameters, including batch size, number of hidden units, learning rate, dropout rates, etc. Inappropriate hyperparameter settings can directly result in suboptimal performance, overfitting, underfitting, or slow and unstable convergence. Traditional hyperparameter tuning methods such as manual tuning, grid search, and random search still remain used but suffer from high computational costs and limited efficiency in exploring complex search spaces, especially contiguous intervals. To overcome these limitations, optimisation algorithms have increasingly been adopted for hyperparameter tuning in machine learning models. Accordingly, the development of optimisation algorithms has emerged as a growing important research topic in financial mathematics (Liu et al., 2017).

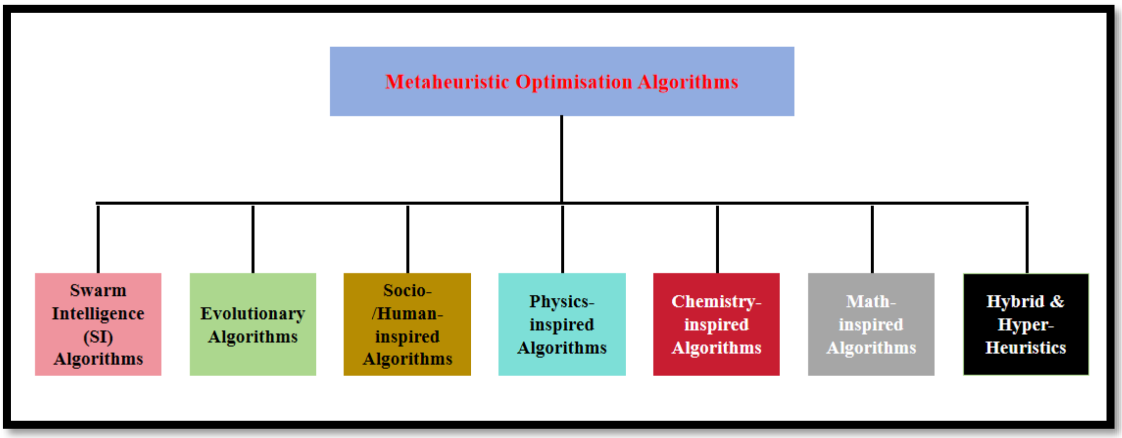

Figure1: The proposed classification of Metaheuristic Algorithms based on the source ofinspiration

Stochastic optimisation algorithms can be classified into Heuristic Algorithms (HAs) and Metaheuristic Algorithms (MAs) (Houssein et al., 2024). Among these, metaheuristic algorithms, as an innovative and powerful optimisation approach, have attracted widespread attention in recent years. These algorithms, inspired by collective behavior and mechanisms observed in nature, implement stochastic optimisation procedures that simulate such behaviors and processes, thereby enabling effective solutions to complex optimisation problems. Therefore, these algorithms can be categorised based on their sources of inspiration, which is one of main taxonomies for metaheuristic algorithms. Figure1clearly presents this classification of metaheuristic algorithms. For example, swarm intelligence algorithms take the inspiration from mimicking the natural behavior of mammals that live in flocks, or herds seeking to hunt for prey or find for food. Examples include particle swarm optimisation (PSO), whose principles follow the cooperative movement of animals such as fish and bird (Bharti et al., 2023; Chegini et al., 2018); the gray wolf optimiser (GWO) algorithm, which is designed on the basis of the hierarchical structure and social behaviour of gray wolves during hunting (Hosseini et al., 2025); the butterfly optimisation algorithm (BOA), which mimics the behaviour of butterflies as they forage for food and search for mating pairs (Sharma & Saha, 2021; Sharma, Chakraborty, et al., 2022; Sharma, Saha, et al., 2022; Sharma, T. K., 2021); the black widow optimisation (BWO) algorithm, which is based on the mating behaviour of black widow spiders (Kushwah & Agrawal, 2024); the sand cat swarm optimisation (SCSO) algorithm, which optimises model parameters by simulating the unique predatory behaviour of sand cat swarms (Han et al., 2025); the whale optimisation algorithm (WOA), a bioheuristic algorithm that seeks the global optimal solution by simulating the social behaviour and feeding strategies of humpback whales (Liu et al., 2025b). Evolutionary algorithm, derived from Darwin's theory of biological evolution, simulate evolutionary processes such as recombination, mutation, crossover, and selection. The prominent algorithms in this class are the genetic algorithm (GA), based on the principle of “survival of the fittest” (Priya & Francis, 2025; Zhong et al., 2018); and the differential evolution (DE), which is a population-based adaptive global optimisation algorithm for solving real number optimisation problems (Vijayaprabakaran & Sathiyamurthy, 2022; Xie et al., 2024). Physics-based algorithms draw upon principles from mechanics, thermodynamics, and electromagnetism. Examples include the gravitational search algorithm (GSA), which is inspired by the physical law of gravity (Naheliya et al., 2024); and the simulated annealing (SA), which is based on the thermodynamic annealing process (Poursaeid et al., 2025). Additionally, there are heuristic algorithms that are based on mathematical theorems, concepts and rules, which adopt optimisation strategies derived from mathematics. For example, the sine cosine algorithm (SCA) utilises sine and cosine functions to guide the search process (Chegini et al., 2018; Mirjalili, 2016; Sharma & Saha, 2021; Sharma, Saha, et al., 2022). These metaheuristic algorithms have been widely applied with distinct characteristics and have been frequently integrated to tackle the diverse fields of optimisation problems, including function optimisation, path planning, feature selection (Liu et al., 2025a), multi-objective or multi-strategy optimisation, neural networks, image processing and many other additional domains.

For example, Pan et al. (2022) proposed a combination model based on genetic algorithm and a multivariable Bi-LSTM model to forecast the sales of filling stations. The results clearly illustrated that, relative to the BP neural network, XGBOOST, exponential smoothing, ARIMA, ANN, different LSTM structures, and Bi-LSTM models, the hybrid GA-Bi-LSTM model achieved the greatest predictive performance, attaining the highest accuracy of 89%. Peng et al. (2018) employed the hybrid model based on the LSTM model and a differential evolution algorithm to predict electricity prices. The results clearly stated that, compared with the rigid regression, equally weighted, inverse of the root mean squared errors, constrained least squares, SVR, SVM, ARIMA, ANN, RNN, BP neural network, hybrid ARIMA-ANN, and hybrid DE-BP neural network models, the hybrid DE-LSTM model achieved the best forecasting performance, delivering the lowest MAE, RMSE, and MAPE. Yang, Ss., et al. (2022) employed Savitzky-Golay filtering to filter the slope monitoring data. Then, they utilised the sparrow search algorithm (SSA) to optimise the hyperparameters of the LSTM neural network for high-accuracy displacement prediction and compared the resulting SSA-LSTM model with the unoptimised LSTM. The results indicate obvious improvements at three of the four monitoring points, with the exception of WY1-2. The gains were most pronounced at WY1-1, where MAE decreased by 2.401, RMSE decreased by 2.086, and R2 increased by 0.199. Moreover, excluding WY1-2, the R2 values at the remaining monitoring points exceeded 0.9. Therefore, these findings clearly suggest that the SSA optimised LSTM model delivers superior displacement prediction performance relative to the LSTM model. Xu et al. (2022) proposed the hybrid model based on a LSTM model and the PSO algorithm. This study utilised the PSO algorithm to optimise the hyperparameters of LSTM model, which improves the ability to learn data sequence features. The results show that the PSO-LSTM model significantly outperforms the M-EIES, ANN, PSO-ANN, and LSTM models at all stations in the watersheds, indicating its potential for application in extreme event prediction. Hu et al. (2021) proposed a BOA incorporating a modified population selection strategy and adaptive sensory modalities. The results of the simulation experiment indicate that, relative to the original BOA as well as the Seagull Optimisation Algorithm and Tunicate Swarm Algorithm, the proposed method achieved improved accuracy and reliability. In addition, comparisons with a BP neural network, BOA-SVM, and KCV-SVM suggest that the proposed hybrid model produced more competitive predictive performance than the alternative approaches. Sharma and Saha (2021) developed a hybrid BOA-SCA algorithm (BOSCA) to improve and stabilise the global exploration and local exploitation ability. In comparative experiments against several popular metaheuristic algorithms, including SCA, GA, DE, BOA, PSO, cockoo search (CS), firefly algorithm (FA), monarch butterfly optimisation (MBO), artificial bee colony (ABC), the BOSCA significantly outperforms than the other compared algorithms on two real-life optimisation problems. Wan et al. (2024) proposed a novel hybrid framework, termed BWO-BiLSTM-ATT, which synergistically integrates a Bi-LSTM network, a self-attention mechanism, and the Beluga Whale Optimisation (BWO) algorithm for accurate wind power forecasting. They evaluated and compared the predictive performance of multiple approaches, including Bi-LSTM, CNN-Bi-LSTM, CNN-Bi-LSTM-ATT, VMD-Bi-LSTM-ATT, Bi-LSTM-ATT, SVR, and RNN, by forecasting the power prediction of four wind turbines. These results clearly illustrate that the Bi-LSTM-ATT model consistently achieved superior performance in terms of various performance metrics, including R-squared, MSE, RMSE, Median Absolute Error (MedAE), MAE, and Explained Variance (EV), demonstrating its effectiveness in capturing intricate temporal dynamics and complex patterns in wind power data. Meanwhile, when comparing Bi-LSTM, Bi-LSTM-ATT, and BWO-BiLSTM-ATT, the proposed framework also attained the highest EV scores of 0.8970 (0.8814, 0.9084), 0.9064 (0.9015, 0.9119), and 0.9108 (0.9094, 0.9131), respectively. These findings show that integrating the self-attention mechanism and the BWO algorithm further enhances its predictive accuracy and generalisation capabilities. Omarzadeh and Hosseini (2025) proposed a novel hybrid approach that combines convolutional neural networks (CNN) and long short-term memory (LSTM) networks, with hyperparameter optimisation guided by the Grey Wolf Optimiser (GWO), along with Glove word embedding, to detect and classify offensive comments on Instagram. The experimental results demonstrated that the proposed GWO-CNN-LSTM hybrid model outperformed baseline models, including standalone LSTM, CNN, and GWO-LSTM, achieving an accuracy of 88.60%, precision of 88.59%, recall of 88.60%, and the F1-score of 88.49%. Overall, these findings suggest that integrating GWO with the CNN-LSTM architecture can significantly enhance model predictive performance.

The Black-winged Kite Algorithm (BKA) application

With continued scientific progress, metaheuristic optimisation algorithms also have evolved to meet the growing demand for faster and more effective hyperparameter tuning when models are employed in increasingly complex real-life forecasting tasks. In particular, bio-inspired optimisation algorithms have attracted considerable attention in recent years. Notably, Wang et al. (2024) were the first to propose a newly designed swarm-intelligence algorithm, named the Black-winged Kite Algorithm (BKA), inspired by the hunting behaviour and migratory strategy of black-winged kites. In addition, this study incorporates a cauchy mutation strategy to help the proposed algorithm escape local optima and to increase the probability of discovering better solutions in the global search space. These properties have contributed to BKA becoming an active topic of recent methodological development.Table 1provides a summary of the experimental comparison results reported in recent studies on improved Black-winged Kite algorithms.For example, Zhou et al. (2024) proposed a hybrid algorithm that integrates BKA with the Sine-Cosine Algorithm (SCA). By introducing a sine-cosine guidance mechanism, the proposed SCBKA Algorithm enhances search efficiency and solution accuracy. These findings clearly illustrate that the SCBKA algorithm substantially outperforms the basic black-winged kite optimisation algorithm (BKA) as well as several benchmark metaheuristics, including the whale optimisation algorithm (WOA), teaching optimisation algorithm (TSA), particle swarm optimisation (PSO), the novel randomised population-based optimisation algorithm (NRBO), and the coati optimisation algorithm (COA) in the field of function optimisation and shows a significant advantage in the field of function optimisation. Zhu et al. (2025) addressed the premature convergence problem of the Black-Winged Kite Algorithm (BKA) in high-dimensional optimisation issues by proposing an enhanced hybrid algorithm termed BKAPI. In this hybrid algorithm, BKA contributes dynamic global exploration via its hovering and dive attack strategies, whereas particle swarm optimisation (PSO) strengthens local exploitation through its velocity-based search mechanism. Moreover, while PSO facilitates efficient local refinement, Differential Evolution (DE) further introduces a differential mutation strategy to maintain population diversity and to reduce the risk of premature convergence. Overall, this hybrid algorithm aims to achieve a more balanced exploration-exploitation trade-off, while overcoming BKA's sensitivity to parameter settings and insufficient local search capabilities. To validate its effectiveness, the proposed algorithm was rigorously evaluated against seven intelligent optimisation methods, namely BKA, PSO, Genetic Algorithm (GA), Cuckoo Optimisation Algorithm (COA), Beluga Whale Optimisation (BWO), Adaptive Optics (AO), and the Starfish Optimisation Algorithm (SFOA), using the CEC 2017 and CEC 2022 benchmark test functions. The empirical results clearly indicate that BKAPI outperformed other advanced algorithms across these benchmarks, suggesting strong optimisation capability and promising application potential. Qu et al. (2025) developed a methane-concentration prediction model for coal-mine roadways by coupling an improved black kite algorithm (IBKA) with an informer-Bi-LSTM model. The traditional BKA algorithm is enhanced by introducing tent chaotic mapping, integrating dynamic convex lens imaging, and adopting a Fraunhofer diffraction search strategy. The research findings clearly indicate that the IBKA algorithm substantially outperforms several benchmark metaheuristics, such as Sailfish Optimiser (SFO), Sparrow Search Algorithm (SSA), SCSO, GWO, and PSO. Moreover, the coupled IBKA-Informer-Bi-LSTM model achieved strong predictive accuracy (MAE = 0.0005971; RMSE = 0.0008005; R^2 = 0.9589) and outperformed the competed models, including LSTM, BiLSTM, Informer, Autoformer, linear regression (LR), and XGBoost. Zhang, H., et al. (2025) proposed a multi-strategy improved black kite algorithm (SWBKA). The proposed algorithm incorporates the latin hypercube sampling to enhance the initial population quality. It then introduces a fused Cauchy-Gaussian disturbance and adaptive spiral position update strategy to balance global exploration and local exploitation during the migration behavior. Finally, a lens imaging reverse learning elite strategy is employed to help the algorithm avoid local optima effectively. The experimental results clearly show that the proposed algorithm achieves an overall average improvement of 47.38% in optimal values compared to BKA. Moreover, when coupled with GRU, the SWBKA-GRU model delivered the best predictive performance relative to the standalone GRU as well as WOA-GRU, (Rime Optimisation Algorithm) RIME-GRU, GWO-GRU, (Subtraction-Average-Based Optimiser) SABO-GRU, and BKA-GRU, achieving a MAPE of 1.321%. Zhang, Bc., et al. (2025) proposed an innovative RUL prediction model based on bidirectional feature fusion. They processed the initial data using variational mode decomposition (VMD) optimised by the black-winged kite algorithm (BKA). Then, the bidirectional features were then extracted via BiTCN and subsequently input into BiGRU. Finally, an attention mechanism was employed to increase emphasis on crucial information. The results illustrate that, across all datasets, the proposed model achieves lower errors than several benchmarks. MAE and RMSE are reduced by 0.006 and 0.009 relative to the BiLSTM-AM model, by 0.008 and 0.014

relative to the CNN-BiGRU-AM model, by 0.01 and 0.014 relative to the CNN-BiLSTM-Bootstrap model, and by 0.01 and 0.015, and 0.002 and 0.005 relative to the DCNN-BiGRU model.

Table 1: Summary of Experimental Comparisons of Improved Black-winged Kite Optimisation Algorithms Reported in the Literature

The improved black-winged kite optimisation algorithms	Ref.	Compared against	Quality metrics
SCBKA	(Zhou et al., 2024)	BKA (Wang et al., 2024),WOA, TSA, PO, NRBO, and COA	CEC 2020 test set (Friedman's rank sum test)
BKAPI	(Zhu et al., 2025)	BKA, PSO, GA,COA, BWO, AO, and SFOA	CEC 2017 and CEC 2022 benchmark test functions (avg and std)
IBKA (Tent chaotic mapping, dynamic convex lens imaging, and a Fraunhofer diffraction search strategy)	(Qu et al., 2025)	SFO, SSA, SCSO, GWO, and PSO	Mean and std
SWBKA	(Zhang, H., et al., 2025)	WOA, GWO, SABO, RIME, and BKA	Mean and std

Conclusion

In the area of financial time-series forecasting, machine learning models have been applied increasingly widely, while hyperparameter optimisation has also become a prominent research topic. To this end, most scholars have devoted considerable attention to the development of metaheuristic optimisation algorithms. Nevertheless, a review of the existing literature indicates that limited research has been conducted on improving the Black-winged Kite Algorithm (BKA) and applying the enhanced algorithm to the hyperparameter optimisation of attention-based LSTM and Bi-LSTM models. Accordingly, this constitutes a promising and worthwhile direction for further investigation.

Author Address:
[1]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
[2]Institute for Mathematical Research, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malysia
[3]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang.

References:

1. Abd Elaziz, M., Dahou, A., Abualigah, L., Yu, Ly., Alshinwan, M., Khasawneh, A. M., & Lu, Sf. (2021). Advanced metaheuristic optimization techniques in applications of deep neural networks: a review. Neural Computing and Applications, 33, 14079-14099.
2. Aggarwal, A., Rani, A., Sharma, P., Kumar, M., Shankar, A., & Alazab, M. (2022). Prediction of Landsliding using Univariate Forecasting Models. Internet Technology Letters, 5(1), e209.
3. Amini, A., & Kalantari, R. (2024). Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning. PLoS ONE. 19(3): e0298426.
4. Bai, Jf., Chen, Zn., Feng, Bl., Xu, B. (2014). Chinese Image Character Recognition Using DNN and Machine Simulated Training Samples. In: Wermter, S., et al. Artificial Neural Networks and Machine Learning-ICANN 2014(ICANN 2014). Lecture Notes in Computer Science, Vol. 8681, pp. 209-216. Springer.
5. Bai, Ym., Xiang, Sy., Cheng, Ff., & Zhao, Js. (2023). A dynamic-inner LSTM prediction method for key alarm variables forecasting in chemical process. Chinese Journal of Chemical Engineering, 55, 266-276.
6. Banerjee, T., Sinha, S., & Choudhury, P. (2022). Long term and short term forecasting of horticultural produce based on the LSTM network model. Applied Intelligence, 52, 9117-9147.
7. Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166.
8. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief, 29, 105340.
9. Bharti, Redhu, P., & Kumar, K. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM. Physica A: Statistical Mechanics and its Applications, 625, 129001.
10. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.

11. Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day Inc., San Francisco, Calif.
12. Cao, Q., Ewing, B. T., & Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. European Journal of Operational Research, 221(1), 148-154.
13. Chegini, S. N., Bagheri, A., & Najafi, F. (2018). PSOSCALF: A new hybrid PSO based on Sine Cosine Algorithm and Levy flight for solving optimization problems. Applied Soft Computing, 73, 697-726.
14. Coşkun, M., Yıldırım, Ö., Uçar, A., & Demır, Y. (2017). AN OVERVIEW OF POPULAR DEEP LEARNING METHODS. European Journal of Technique (EJT), 7(2), 165-176.
15. Csáji, B. C. (2001). Approximation with Artificial Neural Networks. Faculty of Sciences, Etvs Lornd University, Hungary, MSc thesis, 24(48), 7.
16. Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems, 2, 303-314.
17. Dabral, P. P., & Murry, M. Z. (2017). Modelling and Forecasting of Rainfall Time Series Using SARIMA. Environmental Processes, 4, 399-419.
18. Dadhich, M., Pahwa, M. S., Jain, V., & Doshi, R. (2021). Predictive Models for Stock Market Index Using Stochastic Time Series ARIMA Modeling in Emerging Economy. In: Manik, G., Kalia, S., Sahoo, S.K., Sharma, T.K., Verma, O.P. (eds). Advances in Mechanical Engineering, pp. 281-290. Springer.
19. Duan, Gh., Su, Yw., & Fu, J. (2023). Landslide displacement prediction based on multivariate LSTM model. International Journal of Environmental Research and Public Health, 20(2), 1167.
20. Dwijayanti, S., Yamamori, K., & Miyoshi, M. (2018). Enhancement of speech dynamics for voice activity detection using DNN. EURASIP Journal on Audio, Speech, and Music Processing, 2018, Article 10.
21. El Makkaoui, M., Dalli, A., & Elbaamrani, K. (2024). A Comparative Study of RNN and DNN for Climate Prediction. In 2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST). IEEE, pp. 1-6.
22. Franses, P. H., & Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. Journal of Econometrics, 81(1), 273-280.
23. Gandhi, C., Sarangi, P. K., Saxena, M., & Sahoo, A. K. (2023). SMS Spam Detection Using Deep Learning Techniques: A Comparative Analysis of DNN Vs LSTM Vs Bi-LSTM. In 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). IEEE, pp. 189-194.
24. Gorr, W. L. (1994). Research prospective on neural network forecasting. International Journal of Forecasting, 10(1), 1-4.

25. Grubb, H., & Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. International Journal of Forecasting, 17(1), 71-82.
26. Haider, S. A., Naqvi, S. R., Akram, T., Umar, G. A., Shahzad, A., Sial, M. R., Khaliq, S., & Kamran, M. (2019). LSTM Neural Network Based Forecasting Model for Wheat Production in Pakistan. Agronomy-Basel, 9(2), 72.
27. Han, Ht., Peng, Js., Ma, J., Liu, H., & Liu, Sl. (2025). Research on Load Forecasting Prediction Model Based on Modified Sand Cat Swarm Optimization and SelfAttention TCN. Symmetry, 17(8), 1270.
28. Hasib, S. A., Gulzar, M. M., Shakoor, A., Habib, S., & Murtaza, A. F. (2024). Optimizing electric vehicle driving range prediction using deep learning: A deep neural network (DNN) approach. Results in Engineering, 24, 103630.
29. Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. Computers & Industrial Engineering, 35(1-2), 213-216.
30. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
31. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. Neural Networks, 2(5), 359-366.
32. Hosseini, A., Najafi, M., & Morshedy, A. H. (2025). Hybrid ANN-GWO Algorithm for Improving Methane Drainage Efficiency in Cross-Measure Borehole in Underground Coal Mines. International Journal of Energy Research, 2025, Article 1326901.
33. Houssein, E. H., Saeed, M. K., Hu, G., & Al-Sayed, M. M. (2024). Metaheuristics for Solving Global and Engineering Optimization Problems: Review, Applications, Open Issues and Challenges. Archives of Computational Methods in Engineering, 31, 4485-4519.
34. Hu, K., Jiang, H., Ji, C. G., & Pan, Z. (2021). A modified butterfly optimization algorithm: An adaptive algorithm for global optimization and the support vector machine. Expert Systems, 38, 12642.
35. Hu, Xj., Liu, T., Hao, Xt., & Lin, Cx. (2022). Attention-based Conv-LSTM and Bi-LSTM networks for large-scale traffic speed prediction. The Journal of Supercomputing, 78, 12686-12709.
36. Kang, Q., Yu, Dx., Cheong, K. H., & Wang, Z. (2024). Deterministic convergence analysis for regularized long short-term memory and its application to regression and multi-classification problems. Engineering Applications of Artificial Intelligence, 133, Article 108444.
37. Kim, J., & Moon, N. (2019). BiLSTM model based on multivariate time series data in multiple field for forecasting trading area. Journal of Ambient Intelligence and

Humanized Computing.

38. Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review, 7, Article 21.
39. Kushwah, V., & Agrawal, P. (2024). Decomposition techniques and long short term memory model with black widow optimization for stock price prediction. Multimedia Tools and Applications, 83, 37453-37481.
40. Liang, Y., Xiao, T., Hu, C., Ren, Sc., & Zeng, L. (2022). Prediction of Landslide Displacement Based on Long Term Monitoring Data and LSTM Network. Journal of Signal Processing, 38(1), 19-27.
41. Liu, Dp., Wang, X., Cai, Y., Liu, Zh., & Liu, Z.-J. (2020). A Novel Framework of Real-Time Regional Collision Risk Prediction Based on the RNN Approach. Journal of Marine Science and Engineering, 8(3), 224.
42. Liu, Jc., Bi, Cg., Chen, Hl., Heidari, A. A., & Chen, H. (2025a). Triangular-based sine cosine algorithm for global search and feature selection. Scientific Reports 15, Article 12992.
43. Liu, Wb., Wang, Zd., Liu, Xh., Zeng, Ny., Liu, Yr., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
44. Liu, Zy., Li, Xb., Lu, Zg., & Meng, Xh. (2025b). IWOA-RNN: An improved whale optimization algorithm with recurrent neural networks for traffic flow prediction. Alexandria Engineering Journal, 117, 563-576.
45. Lou, H.-R., Wang, X., Gao, Y., & Zeng, Q. (2022). Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China. BMC Public Health, 22, Article 2167.
46. Mao, Ls., Huang, Y., Zhang, Xf., Li, Sj., & Huang, Xn. (2022). ARIMA model forecasting analysis of the prices of multiple vegetables under the impact of the COVID-19. PLoS ONE 17(7), e0271594.
47. Mardjo, A., & Choksuchat, C. (2024). HyBiLSTM: Multivariate Bitcoin Price Forecasting Using Hybrid Time-Series Models With Bidirectional LSTM. IEEE Access, 12, 50792-50808.
48. Mirjalili, S. (2016). SCA: A Sine Cosine Algorithm for solving optimization problems. Knowledge-Based Systems, 96, 120-133.
49. Mo, Hl., & Zhao, Gy. (2024). RIC-CNN: Rotation-Invariant Coordinate Convolutional Neural Network. Pattern Recognition, 146, 109994.
50. Mohammadi, H., & Su, Lx. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics, 32(5), 1001-1008.

51. Naheliya, B., Redhu, P., & Kumar, K. (2024). Bi-directional Long Short Term Memory Neural Network for Short-Term Traffic Speed Prediction Using Gravitational Search Algorithm. International Journal of Intelligent Transportation Systems Research, 22, 316-327.
52. Nyangarika, A., Mikhaylov, A., & Richter, U. H. (2018). Oil Price Factors: Forecasting on the Base of Modified Auto-regressive Integrated Moving Average Model. International Journal of Energy Economics and Policy, 9(1), 149-159.
53. Omarzadeh, H., & Hosseini, M. (2025). Detecting offensive language on instagram with a combined approach of the Gray Wolf algorithm and deep learning networks. Multimedia Systems, 31, Article 401.
54. Pan, S.-Y., Liao, Q., & Liang, Y.-T. (2022). Multivariable sales prediction for filling stations via GA improved BiLSTM. Petroleum Science, 19(5), 2483-2496.
55. Pandey, P. K., Tripura, H., & Pandey, V. (2019). Improving Prediction Accuracy of Rainfall Time Series By Hybrid SARIMA-GARCH Modeling. Natural Resources Research, 28, 1125-1138.
56. Parthiban, T., Ravi, R., & Kalaiselvi, N. (2007). Exploration of artificial neural network [ANN] to predict the electrochemical characteristics of lithium-ion cells. Electrochimica Acta, 53(4), 1877-1882.
57. Pawar, S. R., & Jadhav, S. N. (2018). “A CNN based framework for translation invariant image classification.” 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, pp. 78-82.
58. Peng, L., Liu, S., Liu, R., & Wang, L. (2018). Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy, 162, 1301-1314.
59. Pierre, A. A., Akim, S. A., Semenyo, A. K., & Babiga, B. (2023). Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. Energies, 16(12), 4739.
60. Poursaeid, M., Poursaeed, A.H. & Shabanlou, S. (2025). Water Resources Quality Indicators Monitoring by Nonlinear Programming and Simulated Annealing Optimization with Ensemble Learning Approaches. Water Resources Management, 39, 1073-1087.
61. Priya, C. S., & Francis, F. S. (2025). Optimizing Traffic Speed Prediction Using a Multi-Objective Genetic Algorithm-Enhanced RNN for Intelligent Transportation Systems. IEEE Access, 13, 35688-35706.
62. Qu, H., Shao, Xm., Gao, Hq., Chen, Qj., Guang, Jh., & Liu, C. (2025). A Prediction Model for Methane Concentration in the Buertai Coal Mine Based on Improved Black Kite Algorithm-Informer-Bidirectional Long Short-Term Memory. Processes,

13(1), 205.

63. Ray, S., Das, S. S., Mishra, P., & Khatib, A. M. G. A. (2021). Time Series SARIMA Modelling and Forecasting of Monthly Rainfall and Temperature in the South Asian Countries. Earth Systems and Environment, 5, 531-546.
64. Ray, S., Lama, A., Mishra, P., Biswas, T., Das, S. S., & Gurung, B. (2023). An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. Applied Soft Computing, 149, 110939.
65. Sahu, R., Srivastava, S., & Kumar, R. (2023). Modelling of a Non-Linear Dynamic System Using Long Short-Term Memory. 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 1016-1021.
66. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85-117.
67. Shahani, N. M., Kamran, M., Zheng, Xg., & Liu, Cc. (2022). Predictive modeling of drilling rate index using machine learning approaches: LSTM, simple RNN, and RFA. Petroleum Science and Technology, 40(5), 534-555.
68. Sharma, S., Chakraborty, S., Saha, A. K., Nama, S., & Sahoo, S. K. (2022). mLBOA: A Modified Butterfly Optimization Algorithm with Lagrange Interpolation for Global Optimization. Journal of Bionic Engineering, 19, 1161-1176.
69. Sharma, S., Saha, A. K., Roy, S., Mirjalili, S., & Nama, S. (2022). A mixed sine cosine butterfly optimization algorithm for global optimization and its application. Cluster Computing, 25, 4573-4600.
70. Sharma, S., & Saha, A. K. (2021). BOSCA—A Hybrid Butterfly Optimization Algorithm Modified with Sine Cosine Algorithm. In: Panigrahi, C.R., Pati, B., Mohapatra, P., Buyya, R., Li, KC. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, Vol. 1198, pp. 360-372. Springer, Singapore.
71. Sharma, T. K. (2021). Enhanced butterfly optimization algorithm for reliability optimization problems. Journal of Ambient Intelligence and Humanized Computing. 12, 7595-7619.
72. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) , Orlando, FL, USA, pp. 1394-1401. IEEE.
73. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data)(pp. 3285-3292), Los Angeles, CA, USA.
74. Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit Prediction Using ARIMA,

SARIMA and LSTM Models in Time Series Forecasting: A Comparison. IEEE Access, 10, 124715-124727.

75. Ubal, C., Di-Giorgi, G., Contreras-Reyes, J. E., & Salas, R. (2023). Predicting the Long-Term Dependencies in Time Series Using Recurrent Artificial Neural Networks. Machine Learning and Knowledge Extraction. 5(4), 1340-1358.
76. Usharani, B. (2023). ILF-LSTM: enhanced loss function in LSTM to predict the sea surface temperature. Soft Computing, 27, 13129-13141.
77. Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, 433-441.
78. Vidal, A., & Kristjanpoller, W. (2020). Gold volatility prediction using a CNN-LSTM approach. Expert Systems with Applications, 157, 113481.
79. Vijayaprabakaran, K., & Sathiyamurthy, K. (2022). Towards activation function search for long short-term model network: A differential evolution based approach. Journal of King Saud University - Computer and Information Sciences, 34(6), 2637-2650.
80. Wan, Ap., Peng, S., AL-Bukhaiti, K., Ji, Ys., Ma, Sd., Yao, Fr., & Ao, Lz. (2024). A novel hybrid BWO-BiLSTM-ATT framework for accurate offshore wind power prediction. Ocean Engineering, 312, 119227.
81. Wang, J., Wang, W-c., Hu, X-x., Qiu, L., & Zang H-f. (2024). Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems. Artificial Intelligence Review, 57, Article 98.
82. Wu, C. -H., Lu, C. -C., Ma, Y. -F., & Lu, R. -S. (2018). A new forecasting framework for bitcoin price with LSTM.2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore.
83. Wu, Zh., Pan, Sr., Chen, Fw., Long, Gd., Zhang, Cq., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24.
84. Xie, St., He, Zb., Loh, Y. M., Yang, Y., Liu, Kh., Liu, C., Cheung, C. F., Yu, N., & Wang, Cj. (2024). A novel interpretable predictive model based on ensemble learning and differential evolution algorithm for surface roughness prediction in abrasive water jet polishing. Journal of Intelligent Manufacturing, 35, 2787-2810.
85. Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., & Wang, S. (2022). Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. Journal of Hydrology, 608, 127553.
86. Yang, Eb., Zhang, H., Guo, Xs., Zang, Zn., Liu, Z., & Liu, Yn. (2022). A multivariate multi-step LSTM forecasting model for tuberculosis incidence with model

explanation in Liaoning Province, China. Bmc Infectious Diseases 22, 490.

87. Yang, Ss., Jin, Aj., Nie, W., Liu, C., & Li, Y. (2022). Research on SSA-LSTM-Based Slope Monitoring and Early Warning Model. Sustainability, 14(16), 10246.
88. Yuan, F-C., Lee, C-H., & Chiu, Cc. (2020). Using Market Sentiment Analysis and Genetic Algorithm-Based Least Squares Support Vector Regression to Predict Gold Prices. International Journal of Computational Intelligence Systems, 13, 234-246.
89. Zhang, Bc., Yin, Yg., Li, B., He, Sm., & Song, Jy. (2025). A hybrid algorithm for predicting the remaining service life of hybrid bearings based on bidirectional feature extraction. Measurement, 242, 116152.
90. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.
91. Zhang, H., Liu, Lm., Yang, Yk., & Shao, Wm. (2025). Multi-strategy improved black-winged kite algorithm and its application in short-term load forecasting using gated recurrent unit neural networks. Evolutionary Intelligence, 18, Article 74.
92. Zhong, X., Chen, Ex., Luo, Rq., & Lu, Ys. (2018). Multi-scale fusion dropout optimization algorithm. J. Huazhong Univ. of Sci. & Tech. (Natural Science Edition), 46(9).
93. Zhou, Yq., Wu, X., Liu, Yx., & Jiang, Xh. (2024). BKA optimization algorithm based on sine-cosine guidelines. 2024 4th International Symposium on Computer Technology and Information Science (ISCTIS). IEEE, pp. 480-484.
94. Zhu, Hs., Chen, S., Lu, W., Chen, Kz., Feng, Yl., Xie, Zh., Zhang, Zf., Li, Lf., Ou, Jm., & Chen, Gm. (2022). Study on the influence of meteorological factors on influenza in different regions and predictions based on an LSTM algorithm. Bmc Public Health 22, 2335.
95. Zhu, Xm., Zhang, Js., Jia, Cc., Liu, Y., & Fu, Ms. (2025). A Hybrid Black-Winged Kite Algorithm with PSO and Differential Mutation for Superior Global Optimization and Engineering Applications. Biomimetics (Basel), 10(4), 236.
96. Zraibi, B., Mansouri, M., & Loukili, S. E. (2022). Comparing deep learning methods to predict the remaining useful life of lithium-ion batteries. Materials Today: Proceedings, 62, 6298-6304.