STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–14.
Published online in International Academic Press (www.IAPress.org)

Integrated variable selection with hyperparameter-tuned random forest for modeling exchange rate volatility in Iraq

Abdulrahman Mohammed Ahmed [1], Aws F.A. Jwejatee [1], Zakariya Yahya Algamal [2,3,*]

[1] *Department of Economics, College of Administration and Economics, University of Mosul, Mosul, Iraq*
[2] *Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*
[3] *University of Al-Hamdaniya, Al-Hamdaniya, Iraq*

Abstract This study proposes an integrated random forest framework to forecast exchange rate volatility in Iraq. Traditional econometric models often struggle with the high dimensionality and multicollinearity of macro-financial indicators, which leads to overfitting and weak out-of-sample performance. To address these challenges, the paper develops an integrated sparrow search algorithm–random forest (ISSA-RF) that jointly performs variable selection and hyperparameter tuning. In the first stage, a binary sparrow search algorithm (SSA) selects a parsimonious subset of relevant monetary variables, balancing prediction error and model complexity. In the second stage, a continuous SSA optimizes key random forest hyperparameters, including the number of trees, maximum depth, and node-splitting thresholds. The empirical application uses monthly data from the central bank of Iraq and international sources covering 2004–2024, with a split between training (2004–2020) and testing (2021–2024) samples. Forecasting performance is evaluated using mean absolute error, root mean squared error, coefficient of determination, and directional accuracy, and ISSA-RF is benchmarked against random search, Bayesian optimization, cross-validation, grid search, and SSA-tuned random forests without variable selection. The integrated approach selects six economically meaningful predictors, capturing domestic monetary conditions, foreign exchange interventions, and global monetary and inflation factors. Across both training and testing sets, ISSA-RF consistently delivers the lowest forecast errors, the highest explanatory power, and the strongest directional accuracy, indicating substantial gains in both fit and generalization relative to all competing models. These results suggest that combining metaheuristic-based variable selection with hyperparameter tuning can significantly enhance exchange rate volatility forecasting, providing policymakers and market participants with more reliable tools for risk management and macro-financial surveillance in Iraq.

Keywords Exchange rate volatility, random forest, hyperparameter tuning, sparrow search algorithm, variable selection

AMS 2010 subject classifications 62Fxx, 62P05

DOI: 10.19139/soic-2310-5070-4538

1. Introduction

Volatility of exchange rates is a key challenge for emerging economies, especially if they rely on external income and on geopolitical and macroeconomic uncertainty [1, 40, 41, 42]. A precise depiction is crucial for policymakers, financial institutions, and investors aiming to reduce risks and promote economic stability. The dimensionality of macroeconomic indicators, however, and multicollinearity between the predictors of a model make modelling challenging and result in overfitting and poor predictive ability in traditional econometric models [2, 43, 44]. As a result, forecasting exchange rates has remained an area where there is a high degree of uncertainty and error [3, 46].

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

The critical question has been How can we effectively forecast exchange rate volatility to improve economic decision making and reduce risks in the face of dynamic global financial markets? because it not only summarizes the difficulties that financial analysts, investors and policymakers encounter but also highlights the significance of mastering the intricacies that accompany the dynamics of exchange rates [4, 47, 48]. The exchange rate dynamics in Iraq are shaped by various factors, including changes in oil prices, monetary policy measures, fiscal imbalances, and external shocks, among others.

With the emergence and variety of machine learning methods, it has become possible to address these issues through automated classification systems that can process multi-parameter datasets of large sizes [5, 6]. Random forest has become a popular ensemble learning approach for regression and classification problems, because it can model nonlinear relationships and high order interactions by combining simple decision trees [7, 8, 49, 50, 51]. Random forest combines bootstrap sampling with random feature selection at each split to limit variance and is generally more resistant to overfitting than single decision trees, as it can be used with a variety of mixed-type predictors, allowing for few distribution assumptions when dealing with complex data structures [9, 10, 52]. Finally, the algorithm outputs variable importance measures which can be useful for empirical research to select important variables in the context of high dimensional macro-financial settings. The main reason for its popularity in the field of exchange rates modeling is that it is a set of properties that make it appealing for modeling processes that are likely to be non-linear, regime-dependent, and multi-factorial [11, 12, 53].

Although, random forest has its merits, it has some drawbacks that are more pertinent in the modelling of the volatility of exchange rate. First, the algorithm is often viewed as a black-box model, since averaging over many trees in the ensemble hides what the trees are actually deciding, and makes it hard to interpret the structure of the model in a transparent way [13, 54, 55, 56]. Secondly, the standard variable importance measures are not unbiased in the case of correlated predictors or different scales and cardinalities of predictors, respectively, which may give erroneous conclusions on the relative importance of macroeconomic indicators. Thirdly, random forest can be sensitive to irrelevant or redundant variables, and, in situations where there are many weak variables, the quality of candidate splits can suffer, and predictive performance may be less than desired without some method of variable selection [14]. Finally, hyperparameter selections, such as the number of trees, the tree depth, and the number of variables considered for splitting at each node, can be suboptimal, may have unnecessary computational overhead, or might overfit noisy environments [15]. These restrictions illustrate the need for an integrated framework with explicit variable selection and systematic hyperparameter tuning of the random forest model. This can help to improve the efficiency and reliability of volatility forecasts when both the selection of informative predictors and the calibration of the parameters of the ensemble are taken into account together. Integrated variable selection can reduce dimensionality, reduce the influence of macro-financial indicators on each other, stabilize variable importance assessment, and hyperparameter tuning can control the bias–variance trade-off and optimize the generalization performance of the model.

This study proposes and utilizes an integrated variable selection method together with a hyperparameter optimal random forest model to capture the volatility of the Iraqi exchange rate (IER). The proposed framework sets out to capitalize on the advantages of random forest in capturing nonlinear and interaction effects and explicitly addresses its most important limitations in terms of interpretability, sensitivity to the irrelevant predictors, and the tuning choices, making it a more powerful instrument for policymakers and market participants who are worried about exchange rate risk in Iraq.

2. Data description

In this study, we extracted a monthly dataset from the Central Bank of Iraq (https://www.cbi.iq) showing the market exchange rate of the Iraqi dinar against the US dollar from January 2004 to December 2024, comprising over 240 observations, which constitutes the dependent variable in our study. The training dataset consisted of 204 monthly observations for the period from January 2004 to December 2020, whilst the remaining 48 monthly observations for the period from January 2021 to December 2024 were allocated to the test dataset. Table 1 summarizing some statistical measures. In Figure 1, both the training and the testing data are depicted. In addition, the description of the used monetary variables is listed in Table 2.

Table 1. Statistical measures description

	Training data	Testing data
n	204	48
Mean	1266.775	1492.229
Standard deviation	98.75759	38.96547
Minimum value	1178	1410
Maximum value	1488	1603
Skewness	1.354479	1.207577
Kurtosis	3.3231	4.7373

Table 2. The description of the monetary variables used

Monetary variables ID	Monetary variables description	Source
X1	Policy Rate (%)	https://www.cbi.iq
X2	Money Supply (M2)	https://www.cbi.iq
X3	Inflation Rate (%)	https://www.cbi.iq
X4	Foreign Reserve (US$)	https://www.cbi.iq
X5	Oil Price (US$)	https://www.investing.com
X6	CBI Sales of Foreign Currency (US$)	https://www.cbi.iq
X7	U.S. Dollar Index	https://www.federalreserve.gov
X8	Global inflation rate (inflation rate in US) (%)	https://www.federalreserve.gov
X9	US Federal Funds Rate	https://www.federalreserve.gov

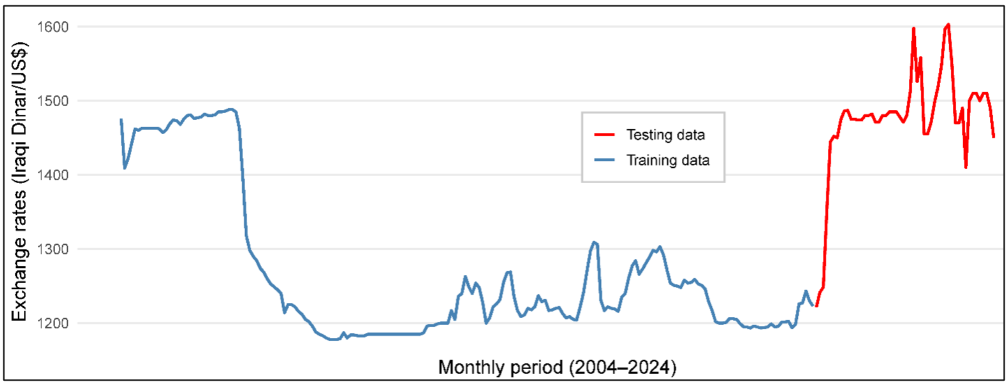

Figure 1. The monthly trends in exchange rate in Iraq (2004–2024) for the training and testing data set.

3. Prediction evaluation criteria

To measure the forecasting performance of the proposed approach, four evaluation criteria, namely, mean absolute error (MAE), root mean squared error (RMSE), direction accuracy (DA), and coefficient of determination (R^2), are selected in this study to verify the prediction accuracy of the proposed model [16, 17]. Their mathematical expressions are listed in Table 3.

Table 3. Prediction evaluation criteria

Evaluation criterion	Mathematical formula	Decision
DA	$\frac{1}{n}\sum_{t=1}^{n} v_t, \quad v_t = \begin{cases} 1, & \text{if } (y_{t+1}-y_t)(\hat{y}_{t+1}-\hat{y}_t) \geq 0 \\ 0, & \text{otherwise} \end{cases}$	Higher value is better
RMSE	$\sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t-\hat{y}_t)^2}$	Lower value is better
R^2	$1-\left[\sum_{t=1}^{n} (y_t-\hat{y}_t)^2 \big/ \sum_{t=1}^{n} (\hat{y}_t-\bar{y}_t)^2\right]$	Larger value near to 1 is better
MAE	$\frac{1}{n}\sum_{t=1}^{n} \lvert y_t-\hat{y}_t\rvert$	Lower value is better

4. Random Forest algorithm

Random forest (RF) algorithm is a popular and potent ensemble machine learning that trains numerous decision trees and uses them to improve classification or regression performance [18]. It has been noted that it is robust, and effective in solving large problems with a high-dimensionality and efficiency over trees by limiting the possibility of overfitting. RF algorithm bases its mechanisms on building the forest of decision trees in training. All trees are grown from randomly sampled subsets of data (using bootstrap sampling) and each split within a tree is done on a random subset of the features rather than all of them. This process adds diversity to trees and hence will serve to improve on the robustness of the model as well as its predictive capabilities [19, 20].

In RF algorithm, the generalization error will converge to be, the greater number of trees. Its performance is related to the strength of individual trees, and the correlation between them [21]. By restricting the size of feature selection made within each split, the resulted induced trees will have a minimal correlation and this will enhance the performance of the ensemble. Gini index is usually adopted to gauge impurity during splitting of nodes. Research has found RF algorithm to be less prone to noise than other algorithms of AdaBoost family such as boosting [22]. RF algorithm has been used in many fields such as predicting academic performance, diagnosis in the medical field, image classification and financial forecasting among others reflecting on its effectiveness and the versatility [23, 24].

5. The proposed improving

RF algorithm is a popular and potent ensemble machine learning. However, standard RF implementations alone do not address the main problems of variable proliferation [25, 26]. The algorithm is fairly noise-tolerant, but if there are many weak or irrelevant predictors in the data, quality of candidate splits may be compromised, computational cost may be increased, and the structure of the data may be hidden. Furthermore, the traditional impurity-based importance measures might be biased when the predictors and (or) variables are correlated, and (or) have different scale or cardinality, which can result in incorrect conclusions for ad-hoc feature selection. The limitations have led to the development of variable selection and random forest modeling as an integrated problem, instead of a series of independent steps. In these integrated systems, information from the ensemble, like the permutation importance, the stability of selecting variables across re-sample, or the conditional importance value, is used systematically to reduce the set of predictors and to select the model specification [27].

The goal of an integrated variable selection with RF is then to take advantage of the benefits of the ensemble and to specifically overcome the drawbacks of the RF in a high-dimensional context [28]. In an iterative filtering and re-estimating process, or by incorporating selection criteria as part of the forest construction, a parsimonious set of predictors, whose predictive ability does not significantly differ from the full set, is obtained, which also improves the interpretability and reduces biases in predictor importance. This is especially useful when, in addition to prediction, substantive interpretations of what covariates are driving the outcome are of interest. From this perspective, the method of integrated variable selection with RF offers a unified method that is both robust and easy to interpret, while also being able to capitalize on modern machine learning techniques [29, 30].

Hyperparameters are those parameters that require defining before implementing an algorithm or before running an algorithm [31]. In the case of Data Science, tuning of hyperparameters is one of the significant stages in the workflow. If done correctly, hyperparameter tuning can turn a worthless model into a model that can be used in production to make decisions [15].

The identification of the appropriate hyperparameters plays a crucial part in the RF prediction accuracy and learning time [32, 57]. However, decision-making of the right combination of hyperparameters is very critical and consumes a lot of time. Five hyperparameters must be tuned in RF algorithm: Eq. (1) Number of trees in the forest, Eq. (2) Minimum number of samples required at a leaf node, Eq. (3) Minimum number of samples needed to split an internal node, Eq. (4) Maximum depth of the tree, and Eq. (5) Number of features considered for the best split [33]. Manual tuning of hyperparameters is time-consuming and at the same time requires more in-depth knowledge of the RF algorithm and associated hyperparameter. As for the manual hyperparameters specification issue, they will need to use the hyperparameter optimization to determine the optimal configuration automatically [34].

Randomized search (RS), Bayesian optimization (BO), cross-validation (CV) and grid search (GS) are the methods that can be employed for the determination of hyper parameters. Previously, it was done with all possible combinations of hyper parameters and the best combination of hyper parameters on the selected criterion was returned. Nonetheless, these are computationally intensive and they still do not cover all the combination of hyperparameters [35].

In recent years, several new nature-based algorithms were suggested by scientists to expand and improve the exploration and exploitation of the existing algorithms. Among its new algorithms, a sparrow search algorithm (SSA) is one of the most popular due to its high performance [36]. The SSA is a swarm intelligence optimization algorithm inspired by the foraging and anti-predation behavior of sparrows. It simulates the intelligent group behavior of sparrows in searching for food and avoiding predators [37]. SSA involves two primary types of sparrows: producers (leaders who search for food) and scroungers (followers who track producers to find food), with certain sparrows acting as scouts or explorers. The algorithm iteratively updates individuals' positions in the search space to find optimal solutions for complex optimization problems [38].

Mathematically, the SSA updates the positions of sparrows based on their roles and environmental conditions using several equations representing their search and avoidance behavior:

First: Position update for producers:

$$y_{k.j}^{t+1} = \begin{cases} y_{k.j}^{t} \times \exp\left(\frac{-k}{\alpha \times iter}\right) & \mathrm{R}_2 < ST \\ y_{k.j}^{t} + qL & \mathrm{R}_2 \geq ST \end{cases} \tag{1}$$

where $y_{k.j}^{t+1}$ indicates the individual position of the k_{th} sparrow on the j_{th} dimension in the $(\mathrm{t}+1)_{th}$ generation; $y_{k.j}^{t}$describes the position of the k_{th} sparrow on the j_{th} dimension in the t_{th} generation ;α is randomly generated from (0,1];iter is the total number of generations; R_2 is the alarm value that belongs to [0,1];ST is the safety threshold generated from [0.5,1]; denotes a random digit following the standard distribution and L expresses the matrix of 1*dim , where dim indicates the dimension of the dataset [39].

Second: Position update for scroungers:

$$y_{k.j}^{t+1} = \begin{cases} q \times \exp\left(\frac{y_{worst}^{t} - x_{k.j}^{t}}{k^2}\right) & \mathrm{k} > \frac{n}{2} \\ y_{best}^{t+1} + \frac{1}{D}\sum_{D=1}^{D}\left(rand \times \left|y_{k.j}^{t} - y_{best}^{t+1}\right|\right) & \mathrm{k} \leq \frac{n}{2} \end{cases} \tag{2}$$

where y_{worst}^{t}indicates the sparrow location having the worst fitness at the t_{th} generation and y_{best}^{t}denotes the location of the producer with the best fitness in the $(\mathrm{t}+1)_{th}$ generation.

Third: Alarm or anti-predation behavior update:

$$y_{k.j}^{t+1} = \begin{cases} y_{best}^{t} + \beta\left|y_{k.j}^{t} - y_{best}^{t}\right| & f_k > f_b \\ y_{k.j}^{t} + h\left(\frac{y_{k.j}^{t} - y_{worst}^{t}}{f_k - f_w + \varepsilon}\right) & f_k = f_b \end{cases} \tag{3}$$

where y_{best}^{t}represents the global optimal position in the t$_{th}$ generation ; β represents the step size control factor ,which follows the normal distribution ;h represents the move direction of sparrow individuals ,which is generated from[-1,1],and it is also a factor of the step size control ; ε is constant used to avoid division be zero errors ; $f_k$indicates the fitness of the k$_{th}$ sparrow; and the best and worst global are represented by f_b and f_w respectively.

The SSA algorithm iteratively updates sparrow positions by these rules, balancing exploration (searching new areas) and exploitation (refining solutions near the current best). The fitness of each individual is evaluated at each step, and individuals move towards better positions based on fitness values to converge to optimal or near-optimal solutions. SSA has demonstrated advantages such as simple structure, fewer parameters, quick convergence, and strong global search capabilities, making it suitable for continuous and discrete optimization problems, including variable selection, function optimization, and engineering.

The proposed approach, integrated variable selection with hyperparameter-tuned random forest using SSA algorithm (ISSA-RF), consists of two basic phases. In the first stage, the variable selection was used through the binary SSA algorithm to filter the redundant variables that are not important in the prediction process. The fitness function used in the binary SSA algorithm is as follows:

$$Fitness = z_1 \times \text{MAE} + z_2 \times \left(\frac{vq}{vp - vq}\right), \tag{4}$$

where MAE is the mean absolute error which is the defined in Table 3, vq is the length of selected variables subset, vp is the total number of variables in the dataset, and z_1 and z_2 are two parameters corresponding to the weight of prediction accuracy and variable selection quality, $z_1 \in [0, 1]$ and $z_2 = 1 - z_1$.

The second phase of the proposed approach focuses on optimizing the hyperparameters of RF, specifically the five hyperparameters: Number of trees in the forest (N$_{trees}$), Minimum number of samples required at a leaf node (MN$_{leaf}$), Minimum number of samples needed to split an internal node (MN$_{internal}$), Maximum depth of the tree (M$_{tree}$), and Number of features considered for the best split (N$_{features}$). These hyperparameters are obtained through the subset of the optimal properties obtained in the first stage based on the variable selection method. The selection of RF hyperparameters directly affects the prediction performance. A representation of the purpose of ISSA-RF is shown in Figure 2. The flowchart of the proposed ISSA-RF framework is presented in Figure 3.

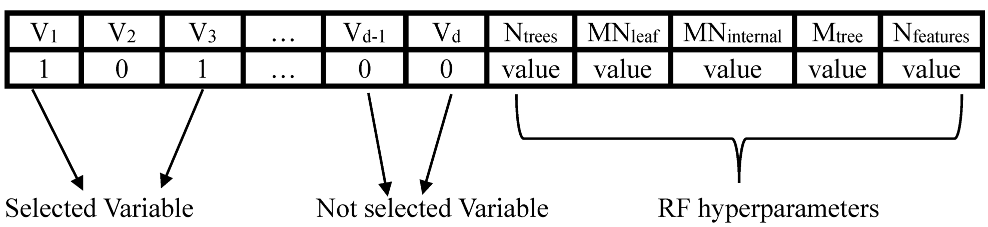

Figure 2. A representation of the ISSA-RF algorithm.

Consequently, our proposed improving is as:

Step 1: The number of sparrows, $N_{sparrow}$, is set to 30 and the maximum number of iterations is T= 1000.

Step 2: The positions of each sparrow are randomly specified. For the N$_{trees}$, the position is randomly generated from uniform distribution with 8 and 100. While for the MN$_{leaf}$ and for the MN$_{internal}$, the position for each one is randomly generated from uniform distribution within the range [1,40] and range [2,30], respectively. Further, the position of the M$_{tree}$ is randomly generated from uniform distribution with 1 and 300. Finally, the N$_{features}$ taken for the best split is considered depending on the number of lags of the series. Related to the variable selection, the positions are randomly assigned randomly with the uniform distribution using 0 and 1.

Step 3: The fitness function is defined as in Eq. (4).

Step 4: The positions of the sparrows are updated using Eq. (1), Eq. (2), and Eq. (3), respectively.

Step 5: Steps 3 and 4 are repeated until a T is reached.

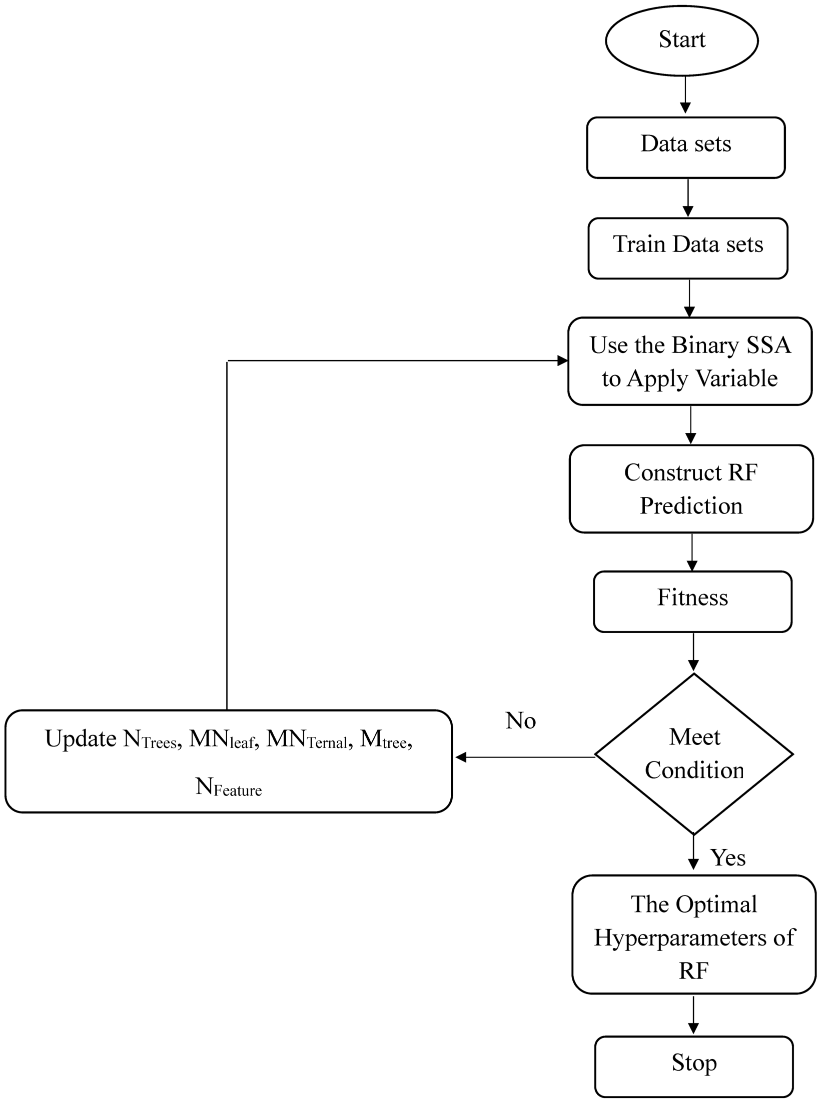

Figure 3. The architecture of the proposed ISSA-RF algorithm.

6. Results and discussion

To enhance the performance for predicting exchange rate volatility in Iraq, the proposed algorithm, ISSA-RF, is compared with the following:

1. Randomized search for tuning RF hyperparameters without variable selection (RF-RS).
2. Bayesian optimization for tuning RF hyperparameters without variable selection (RF-BO).
3. Cross-validation for tuning RF hyperparameters without variable selection (RF-CV).
4. Grid search for tuning RF hyperparameters without variable selection (RF-GS).
5. SSA for tuning RF hyperparameters without variable selection (RF-SSA).

The prediction performance using the evaluation criteria were listed in Tables 4 and 5 for both training and testing data.

Table 4. The prediction performance for exchange rate volatility in Iraq based on training set

Method	DA	RMSE	R^2	MAE
RF-RS	0.455	6.412	0.748	6.294
RF-BO	0.676	5.321	0.811	5.266
RF-CV	0.663	5.103	0.802	5.059
RF-GS	0.461	6.429	0.744	6.387
RF-SSA	0.764	2.188	0.917	2.139
ISSA-RF	0.812	1.347	0.951	1.116

Table 5. The prediction performance for exchange rate volatility in Iraq based on testing set

Method	DA	RMSE	R^2	MAE
RF-RS	0.411	6.361	0.704	6.241
RF-BO	0.632	5.273	0.767	5.213
RF-CV	0.619	5.052	0.758	5.006
RF-GS	0.417	6.378	0.702	6.334
RF-SSA	0.721	2.137	0.873	2.086
ISSA-RF	0.768	1.296	0.907	1.063

Table 4 reports the in-sample (training) performance of six random forest-based models for predicting Iraqi exchange rate volatility using four criteria. In this table, the proposed ISSA-RF clearly dominates all competing methods. It attains the highest DA with 0.812, indicating that it most accurately captures the direction of changes in the exchange rate, and it achieves the lowest RMSE with 1.347 and MAE with 1.116, meaning its point forecasts are the closest to the actual values. At the same time, it produces the highest R^2 with 0.951, showing that it explains a very large proportion of the variability in the training data. Compared with the other hyperparameter tuning strategies, RF-SSA already yields a major improvement in all metrics, but ISSA-RF further refines this by eliminating redundant predictors and fine-tuning the RF structure, leading to a more accurate and parsimonious model rather than just a more complex one.

Further, when we move from these classical methods to RF-SSA, there is a substantial jump in performance: RMSE drops from around 5 to 2.188, MAE from around 5 to 2.139, R^2 rises to 0.917, and DA increases to 0.764. This indicates that using SSA for hyperparameter tuning alone already yields a much more accurate random forest than BO, CV, RS, or GS. Finally, comparing RF-SSA with the proposed ISSA-RF isolates the added value of the integrated variable selection step. ISSA-RF further reduces RMSE from 2.188 to 1.347 and MAE from 2.139 to 1.116, increases DA from 0.764 to 0.812, and pushes R^2 from 0.917 to 0.951. In other words, relative to the best alternative (RF-SSA), ISSA-RF improves directional accuracy by about 4.8% and cuts the remaining error by roughly 35–45%, demonstrating that combining SSA-based tuning with feature selection produces a clearly more efficient and robust model than tuning alone.

Table 5 presents the out-of-sample (testing) performance for the same set of models over the more recent period. This table is crucial because it shows how well each method generalizes to unseen data, which is the main concern of referees. The same pattern persists, where the conventional tuning methods (RF-RS, RF-BO, RF-CV, RF-GS) perform moderately, with DA values mostly between about 0.41 and 0.63 and error measures (RMSE and MAE) around 5–6. Introducing SSA for hyperparameter tuning substantially improves the forecasts (RF-SSA), reducing the errors and increasing both DA and R^2. However, ISSA-RF again provides the best results among all methods, with the highest DA 0.768, the lowest RMSE 1.296 and MAE 1.063, and the highest R^2 0.907 on the test set. This indicates that the gains observed in the training sample are not due to overfitting: the model that combines SSA-based tuning with integrated variable selection genuinely improves predictive power and directional forecasting in a robust way. Economically, this means that the proposed model not only predicts the level of exchange rate volatility more accurately but also offers more reliable signals about whether volatility will increase or decrease, which is particularly valuable for policymakers, investors, and risk managers dealing with the Iraqi exchange rate.

In terms of variable selection, the ISSA-RF selected six important monetary variables out of nine monetary variables. They are: policy rate (X1), money supply (X2), inflation rate (X3), CBI sales of foreign currency (X6),

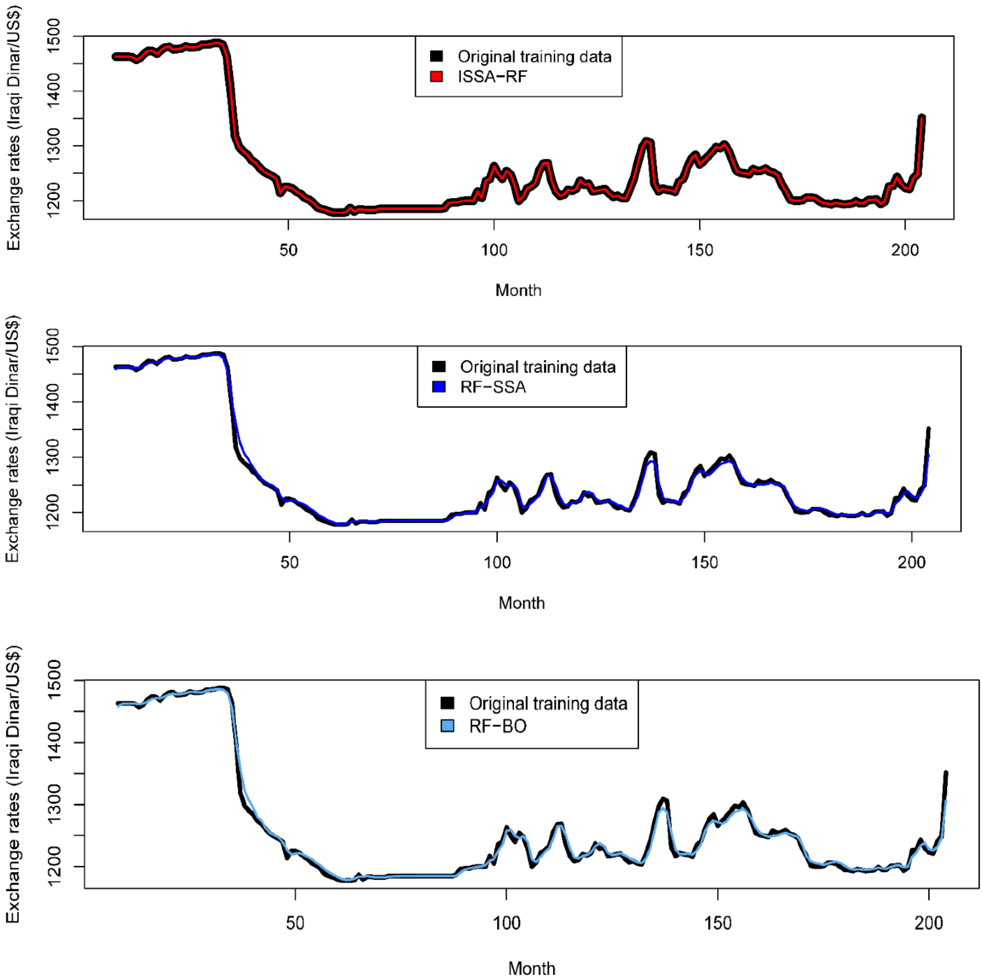

Figure 4. Prediction results in training data for ISSA-RF, RF-SSA, and RF-BO.

global inflation rate (X8), and US federal funds rate (X9). Economically, the fact that ISSA-RF selected those six variables means that, among all the monetary indicators considered, these six carry the most information about future exchange rate volatility in Iraq. In other words, when the algorithm is allowed to drop redundant or weak predictors, it finds that movements in the Iraqi exchange rate are most closely tied to these six variables. The three domestic monetary variables: policy rate (X1), money supply (X2), and inflation rate (X3) capture the stance and credibility of Iraqi monetary policy. When the Central Bank tightens or loosens its policy rate, or when money supply grows faster than real activity, this affects inflation expectations and interest rate differentials the rest of the world, which in turn influence capital flows and demand for foreign currency. Higher domestic inflation (X3) typically erodes the real value of the dinar and increases uncertainty about future prices, which tends to raise exchange rate volatility, especially in an economy heavily dependent on imports and oil revenues. CBI sales of foreign currency (X6) reflect direct FX market interventions by the Central Bank of Iraq. When the CBI increases its dollar sales, it can smooth short-term pressures on the Iraqi dinar, narrow the gap between official and market rates, and temporarily reduce volatility; conversely, reduced sales or constraints on FX supply can amplify volatility when demand for dollars remains strong. The algorithm's choice of X6 as an important predictor therefore indicates that the timing and scale of intervention operations are a key driver of how turbulent the exchange rate becomes.

Moreover, the two external variables, global inflation (X8) and the US federal funds rate (X9), capture the international monetary environment faced by Iraq as a small open, dollar-linked economy. Global inflation shapes world interest rates and risk sentiment, while the US federal funds rate, as the anchor for dollar funding costs, influences global capital flows, the attractiveness of holding dollars versus dinars, and thus the pressure on Iraq's exchange rate regime. When the Fed tightens, for example, dollar assets become more attractive, which can trigger outflows from emerging markets and raise volatility in their exchange rates, especially if domestic conditions are fragile. The selection of X8 and X9 indicating that Iraqi exchange rate volatility is not determined only by local policy, but also by global monetary tightening/loosening cycles and worldwide inflation shocks.

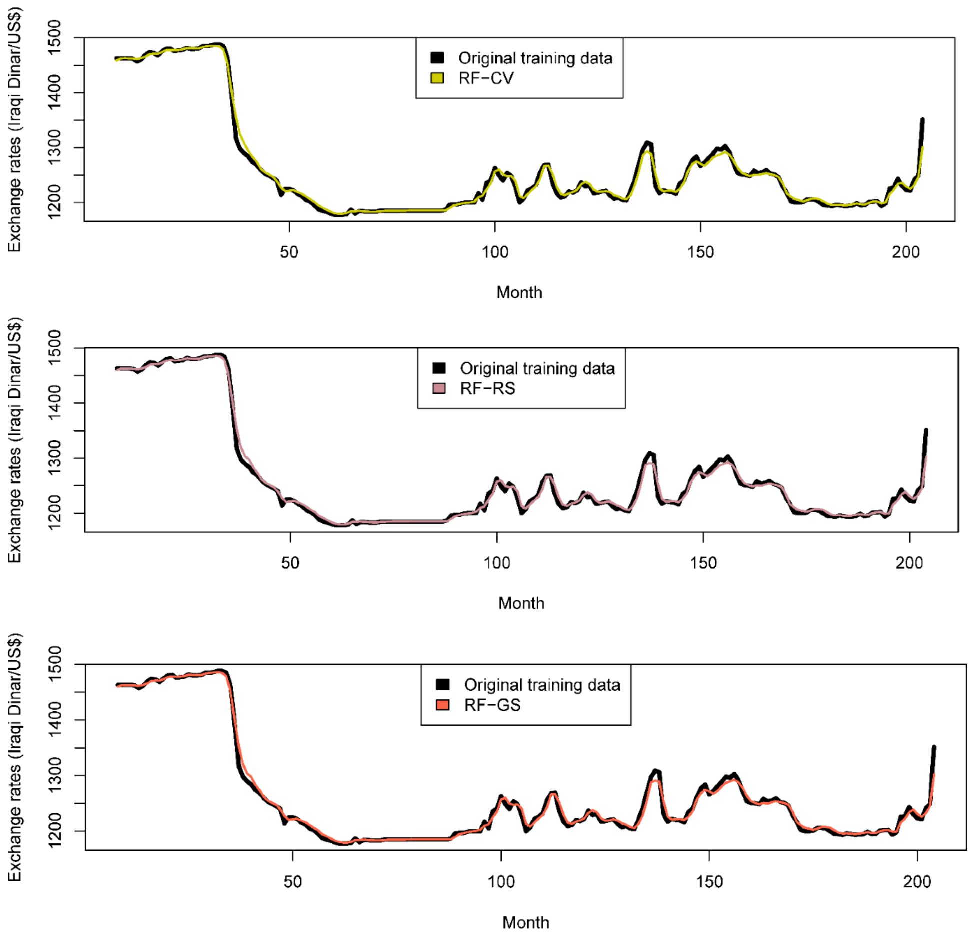

Figure 5. Prediction results in training data for RF-CV, RF-RS, and RF-GS.

Figures 4 and 5 are visual complements to Table 4. These figures show, on the training data, how well the different random forest models track the actual Iraqi exchange rate series over time. Figure 4 focuses on the three best-performing models in terms of training metrics: ISSA-RF, RF-SSA, and RF-BO. Visually, the fitted line from ISSA-RF lies closest to the actual exchange rate path, especially around episodes where the dinar exhibits noticeable changes or volatility spikes. RF-SSA also follows the actual series quite closely but with slightly larger deviations in some periods, while RF-BO shows more visible gaps between predicted and actual values. This graphical pattern is consistent with Table 4, where ISSA-RF has the smallest RMSE and MAE and the highest DA and R^2 among all models.

Additionally, Figure 5 instead displays the prediction results for RF-CV, RF-RS, and RF-GS, which are the more conventional hyperparameter tuning approaches. In this figure, the distance between the fitted lines and the actual exchange rate is generally larger and more systematic. These models tend to under- or over-shoot the actual series in several sub-periods, and they miss turning points more often. This visual impression echoes their weaker numerical performance in Table 4. Economically, the comparison between Figures 4 and 5 reinforces the message that the proposed ISSA-RF framework not only improves summary error metrics but also provides a much tighter fit to the historical path of the exchange rate, capturing both levels and directional movements more accurately than standard RF tuning strategies.

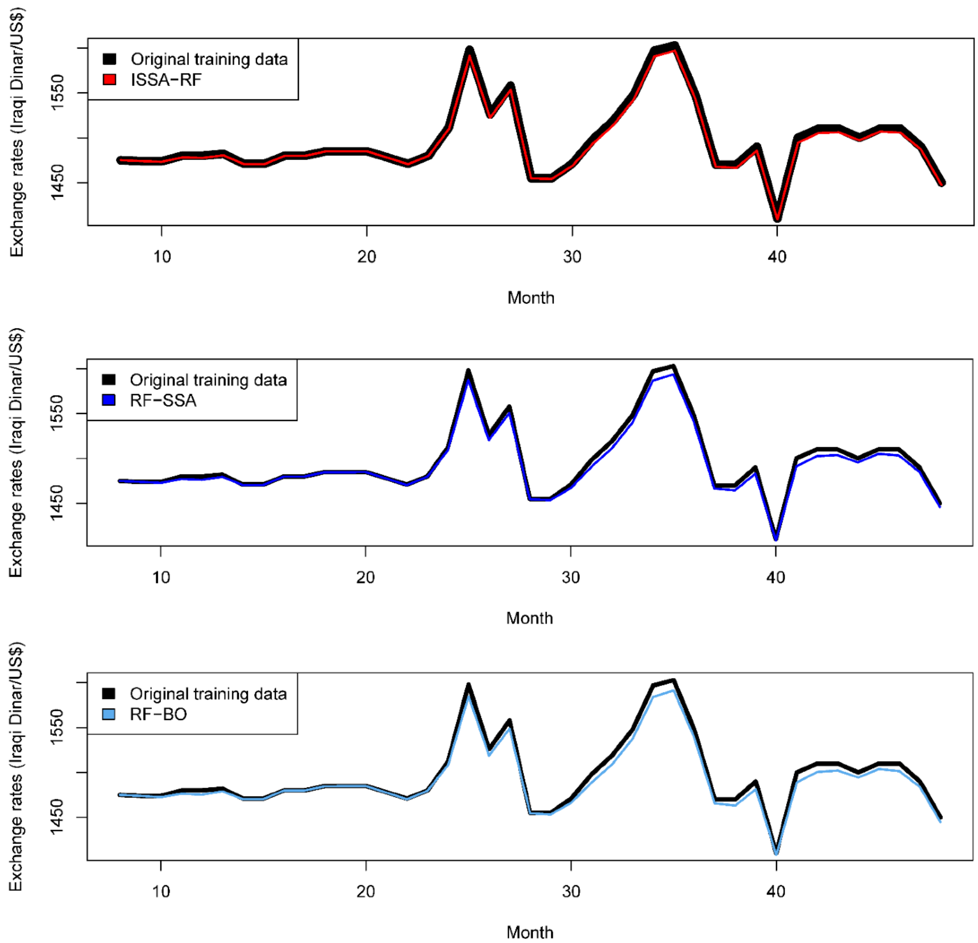

Figure 6. Prediction results in testing data for ISSA-RF, RF-SSA, and RF-BO.

Figures 6 and 7 are the out-of-sample counterparts of Figures 4 and 5. They show how well each model predicts the exchange rate in the testing period (2021–2024). Figure 6 displays the testing predictions for the three strongest models according to Table 5, ISSA-RF (the proposed method), RF-SSA, and RF-BO. In this figure, the ISSA-RF curve lies very close to the actual exchange rate across most months, including during periods where the exchange rate changes more sharply, indicating that it captures both the level and turning points of the series with high accuracy. RF-SSA also tracks the actual path reasonably well but shows larger deviations in some episodes, while RF-BO's prediction line is visibly further from the observed series, missing some of the finer movements. This visual evidence is consistent with the testing metrics in Table 5, where ISSA-RF achieves the smallest RMSE and MAE and the highest DA and R^2, followed by RF-SSA, and then RF-BO. Thus, Figure 6 graphically confirms that the combination of SSA-based tuning and integrated variable selection provides the best generalization to new data.

Further, Figure 7 presents the testing prediction results for RF-CV, RF-RS, and RF-GS, which are the more traditional hyperparameter tuning strategies without SSA or variable selection. In this figure, the gap between the predicted lines and the actual exchange rate is generally wider and more persistent than in Figure 6. These models tend to systematically under- or over-predict the exchange rate in certain sub-periods, and they are less able to follow sudden changes or shifts in the trend. This behavior aligns with their weaker performance in Table 5. Economically, the contrast between Figures 6 and 7 underscores that the proposed ISSA-RF model provides more reliable and stable forecasts in real time, which is crucial for policy and risk management, while models tuned only by RS, CV, or GS tend to produce noisier and less accurate predictions when confronted with new exchange rate conditions.

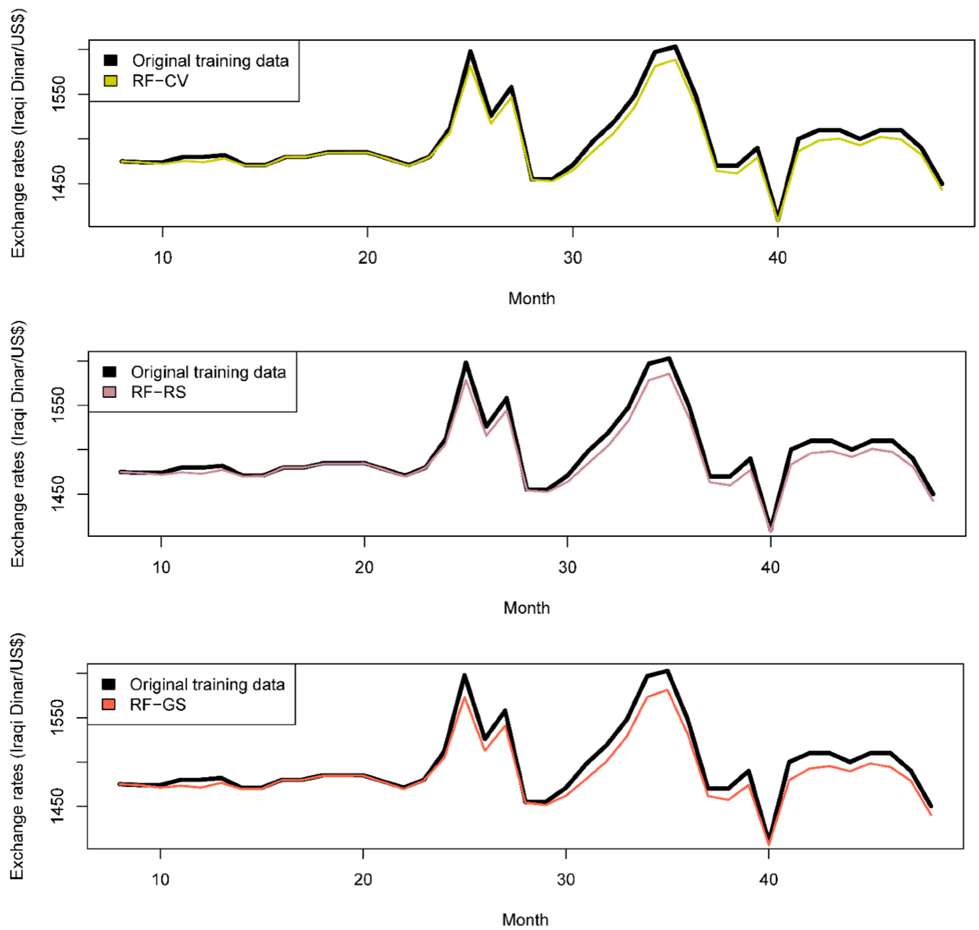

Figure 7. Prediction results in testing data for RF-CV, RF-RS, and RF-GS.

7. Conclusion

This study investigated the volatility of the Iraqi dinar–US dollar exchange rate using an integrated RF framework that combines variable selection with metaheuristic hyperparameter tuning. Based on monthly data from 2004 to 2024, the proposed ISSA-RF model was compared with several benchmark random forest configurations whose hyperparameters were tuned using RS, BO, CV, GS, and a standard SSA without embedded variable selection. The models were evaluated using several criteria. The empirical findings show that ISSA-RF consistently outperforms all alternatives on both the training and testing samples, achieving the lowest MAE and RMSE and the highest R^2 and DA. In particular, relative to the best competing tuned RF, ISSA-RF substantially reduces forecast errors and improves the ability to correctly predict the direction of exchange rate movements, confirming that jointly optimizing feature selection and hyperparameters yields a more accurate and robust forecasting tool. The integrated selection stage identifies six economically meaningful monetary variables as the most relevant predictors, capturing domestic monetary conditions, the Central Bank of Iraq's foreign currency sales, and global inflation and interest rate developments. This parsimonious specification enhances interpretability while preserving, and even improving, predictive performance. Nevertheless, the study has several limitations. First, it focuses on a single country and exchange rate pair, which may limit the generalizability of the results to other currency regimes or institutional settings. Second, only a specific set of macro-financial variables is considered; important political, fiscal, or micro-structure indicators are not explicitly modeled. Third, the analysis is based on a particular ensemble method and a single metaheuristic optimizer; alternative machine learning architectures and optimization strategies may yield further gains. Future research could extend this framework to multi-country or multi-currency settings, incorporate additional macroeconomic and geopolitical variables, and explore deep learning or hybrid econometric–machine learning models. It would also be valuable to compare ISSA-RF with other state-of-the-art variable selection and metaheuristic algorithms, and to embed the forecasts into policy simulation or stress-testing exercises to better quantify the benefits for monetary and exchange rate management in Iraq and similar economies.

REFERENCES

1. D. Umoru, et al., *Forecasting volatilities of currency exchange rates of emerging market economies*, Risk Governance & Control: Financial Markets & Institutions, vol. 15, no. 3, 2025.
2. T. Luo, et al., *Enhancing exchange rate volatility prediction accuracy: Assessing the influence of different indices on the USD/CNY exchange rate*, Finance Research Letters, vol. 58, 2023.
3. O. Abouzaid, and F. Boussedra, *Artificial intelligence and exchange rate forecasting: Assessing predictive accuracy and macroeconomic sensitivity*, Frontiers in Applied Mathematics and Statistics, vol. 11, 2025.
4. P. E. Eniayewu, et al., *Forecasting exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach*, Scientific African, vol. 23, 2024.
5. F. Ding, et al., *Optimization of water quality index models using machine learning approaches*, Water Research, vol. 243, 2023.
6. M. G. Uddin, et al., *Data-driven evolution of water quality models: An in-depth investigation of innovative outlier detection approaches—a case study of Irish Water Quality Index (IEWQI) model*, Water Research, vol. 255, 2024.
7. Z. Nafi Shehab, Z. Y. Algamal, and R. M. Faisal, *Forecasting PM2.5 daily concentration in Baghdad, Iraq based on improving random forest algorithm*, Iraqi National Journal of Earth Science (INJES), vol. 26, no. 1, 2025.
8. M. Arashi, A. F. Lukman, and Z. Y. Algamal, *Liu regression after random forest for prediction and modeling in high dimension*, Journal of Chemometrics, vol. 36, no. 4, 2022.
9. C. Luong, and N. Dokuchaev, *Forecasting of realised volatility with the random forests algorithm*, Journal of Risk and Financial Management, vol. 11, no. 4, 2018.
10. O. Abouzaid, and F. Boussedra, *Artificial intelligence and exchange rate forecasting: Assessing predictive accuracy and macroeconomic sensitivity*, Frontiers in Applied Mathematics and Statistics, vol. 11, 2025.
11. S. Baya, et al., *Modeling and forecasting the real effective exchange rate in Morocco: A comparative analysis by ARIMA, random forest and the dynamic factor model*, in *ICATH 2025*, 2025.
12. A. Olsen, et al., *Forecasting implied volatilities of currency options with machine learning techniques and econometrics models*, International Journal of Data Science and Analytics, vol. 20, no. 2, pp. 1329–1347, 2024.
13. S. Gu, B. Kelly, and D. Xiu, *Empirical asset pricing via machine learning*, The Review of Financial Studies, vol. 33, no. 5, pp. 2223–2273, 2020.
14. J. L. Speiser, et al., *A comparison of random forest variable selection methods for classification prediction modeling*, Expert Systems with Applications, vol. 134, pp. 93–101, 2019.
15. P. Probst, M. N. Wright, and A. L. Boulesteix, *Hyperparameters and tuning strategies for random forest*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 3, p. e1301, 2019.
16. H. Zhang, et al., *Forecasting monthly copper price: A comparative study of various machine learning-based methods*, Resources Policy, vol. 73, 2021.
17. X. Li, et al., *Forecasting the lithium mineral resources prices in China: Evidence with Facebook Prophet (Fb-P) and artificial neural networks (ANN) methods*, Resources Policy, vol. 82, p. 103580, 2023.
18. J. Patel, et al., *A machine learning-based water potability prediction model by using synthetic minority oversampling technique and explainable AI*, Computational Intelligence and Neuroscience, vol. 2022, no. 1, p. 9283293, 2022.
19. P. K. Jena, et al., *Surface water quality assessment by random forest*, Water Practice & Technology, vol. 18, no. 1, pp. 201–214, 2023.
20. D. A. Dewi, et al., *Water quality prediction using random forest algorithm and optimization*, Journal of Applied Data Sciences, vol. 5, no. 3, pp. 1354–1362, 2024.

21. S. Devi, *Random forest advice for water quality prediction in the regions of Kadapa district*, International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 6, pp. 1464–1466, 2019.
22. P. Kamath, et al., *Exploratory data analysis and water potability classification using supervised machine learning algorithms*, Engineering, Technology & Applied Science Research, vol. 15, no. 2, pp. 20898–20903, 2025.
23. M. Yurtsever, and M. Emeç, *Potable water quality prediction using artificial intelligence and machine learning algorithms for better sustainability*, Ege Academic Review, vol. 23, no. 2, pp. 265–278, 2023.
24. A. Lokman, W. Z. W. Ismail, and N. A. A. Aziz, *A review of water quality forecasting and classification using machine learning models and statistical analysis*, Water, vol. 17, no. 15, p. 2243, 2025.
25. J. Ding, et al., *A novel two-stage feature selection method based on random forest and improved genetic algorithm for enhancing classification in machine learning*, Scientific Reports, vol. 15, no. 1, p. 16828, 2025.
26. R. Iranzad, and X. Liu, *A review of random forest-based feature selection methods for data science education and applications*, International Journal of Data Science and Analytics, vol. 20, no. 2, pp. 197–211, 2024.
27. H. Zhang, et al., *A combination of feature selection and random forest techniques to solve a problem related to blast-induced ground vibration*, Applied Sciences, vol. 10, no. 3, 2020.
28. F. Wang, et al., *A hybrid model integrating improved flower pollination algorithm-based feature selection and improved random forest for NO_X emission estimation of coal-fired power plants*, Measurement, vol. 125, pp. 303–312, 2018.
29. M. Rostami, and M. Oussalah, *A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest*, Informatics in Medicine Unlocked, vol. 30, p. 100941, 2022.
30. J. L. Speiser, *A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data*, Journal of Biomedical Informatics, vol. 117, p. 103763, 2021.
31. R. Khalid, and N. Javaid, *A survey on hyperparameters optimization algorithms of forecasting models in smart grid*, Sustainable Cities and Society, vol. 61, p. 102275, 2020.
32. M. Daviran, et al., *A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach*, Computers & Geosciences, vol. 148, p. 104688, 2021.
33. N. Zhu, et al., *Optimization of the random forest hyperparameters for power industrial control systems intrusion detection using an improved grid search algorithm*, Applied Sciences, vol. 12, no. 20, p. 10456, 2022.
34. D.-M. Ge, L.-C. Zhao, and M. Esmaeili-Falak, *Estimation of rapid chloride permeability of SCC using hyperparameters optimized random forest models*, Journal of Sustainable Cement-Based Materials, vol. 12, no. 5, pp. 542–560, 2023.
35. E. Abdulsaed, M. Alabbas, and R. Khudeyer, *Hyperparameter optimization for convolutional neural networks using the salp swarm algorithm*, Informatica, vol. 47, no. 9, 2023.
36. J. Xue, and B. Shen, *A survey on sparrow search algorithms and their applications*, International Journal of Systems Science, vol. 55, no. 4, pp. 814–832, 2024.
37. J. Xue, and B. Shen, *A novel swarm intelligence optimization approach: Sparrow search algorithm*, Systems Science & Control Engineering, vol. 8, no. 1, pp. 22–34, 2020.
38. L. Sun, et al., *BSSFS: Binary sparrow search algorithm for feature selection*, International Journal of Machine Learning and Cybernetics, vol. 14, no. 8, pp. 2633–2657, 2023.
39. L. Lian, *An improved sparrow search algorithm using chaotic opposition-based learning and hybrid updating rules*, Concurrency and Computation: Practice and Experience, vol. 36, no. 14, 2024.
40. Al-Taweel, Y., & Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
41. Algamal, Z. Y., & Lee, M. H. *Applying penalized binary logistic regression with correlation based elastic net for variables selection*, Journal of Modern Applied Statistical Methods, vol. 14, no.(1), p.15, 2015
42. Ewees, A. A., Algamal, Z. Y., Abualigah, L., Al-Qaness, M. A., Yousri, D., Ghoniem, R. M., & Abd Elaziz, M. *A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators*, Mathematics, vol. 10, no.(8), p.1273, 2022
43. Algamal, Z. Y., Lukman, A. F., Abonazel, M. R., & Awwad, F. A. *Performance of the ridge and Liu estimators in the zero-inflated Bell regression model* Scientifica, vol. 1, p. 9503460, 2022.
44. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR modelling using penalized linear regression model with L 1/2-norm*, SAR and QSAR in Environmental Research, vol. 27, no.(9), p.703-719, 2016
45. Shamany, R., Alobaidi, N. N., & Algamal, Z. Y. *A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 12, no.(2), p.453-464, 2019
46. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm*, SAR and QSAR in Environmental Research, vol.31, no.11, p.803-814, 2020.
47. Al-Fakih, A. M., Algamal, Z. Y., Lee, M. H., Aziz, M., & Ali, H. T. M. *QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm*, SAR and QSAR in Environmental Research, vol.30, no.2, p.131-143, 2019.
48. M. Y. Mustafa & Z. Y. Algamal *Neutrosophic inverse power Lindley distribution: A modeling and application for bladder cancer patients* International Journal of Neutrosophic Science, vol. 21, no. 2, pp.216-223, 2023.
49. Qasim, O. S., & Algamal, Z. Y. *A gray wolf algorithm for feature and parameter selection of support vector classification* , International Journal of Computing Science and Mathematics, vol. 13, 93-102, 2021.
50. Salih, A. M., Algamal, Z., & Khaleel, M. A. *A new ridge-type estimator for the gamma regression model*, Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98, 2023.
51. Al-Fakih, A. M., Aziz, M., Abdallah, H. H., Algamal, Z. Y., Lee, M. H., & Maarof, H. *High dimensional QSAR study of mild steel corrosion inhibition in acidic medium by furan derivatives* , International Journal of Electrochemical Science, vol. 10, no. 4, p. 3568-3583, 2015.

52. Algamal, Z. Y., Alhamzawi, R., & Ali, H. T. M. *Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression*, Computers in biology and medicine, vol. 97, p. 145-152, 2018
53. Gatea, A. H., Sadik, N. J., Rasheed, A. A., & Algamal, Z. Y. *Improving SVR Horse Herd Optimization Algorithm for Forecasting Daily Gold Prices* , International Journal on Advanced Science, Engineering Information Technology, 16(2), p.656, 2026.
54. Ibrahim, O.S., & Algamal, Z.Y. *Hybrid V-Support Vector Regression as Statistical Methodology for Forecasting Daily Natural Gas Prices*, Industrial Engineering and Management Systems, vol. 24, no. 4, p. 694-705, 2025.
55. Ibrahim, O.S., & Algamal, Z.Y. *Improving Nonlinear Regression Model Estimation Based on Coati Optimization Algorithm*, Statistics Optimization and Information Computing, vol. 14, no. 2, p. 923-936, 2025.
56. Abdullah, N. A., & Algamal, Z. Y. *). Hybrid Interval Forecasting Model for Iraqi Stock Prices Based on Optimized v Support Vector Regression*, Journal of Mathematics and Modeling in Finance, 6(2), p. 79-97, 2026.
57. Yahya, I. M., & Algamal, Z. Y. *Forecasting Big Data Daily Silver Price Based on Hybridizing Support Vector Regression and Meta-Heuristic Algorithms* Baghdad Science Journal, vol. 22, no. 11, p. 3891-3899, 2025.