International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

PSO-Optimized Two-Level Stacking Ensemble with Linear and Nonlinear Meta-Learners for Numerical Data Imputation

Bilal Ibrahim Maijamaa[1], Salim Ahmad[2], Zaharaddeen Salele Iro[2], Aminu Aliyu Abdullahi[1]
[1]Computer Science Department, Federal University Dutse, Dutse, Nigeria
[2]Information Technology Department, Federal University Dutse, Dutse, Nigeria

Abstract — Missing values are a common challenge in real-world datasets, often reducing the accuracy and reliability of machine learning models. Although numerous machine learning and ensemble-based imputation techniques have been proposed, many rely on single-level stacking architectures and do not explicitly model both linear and nonlinear relationships during prediction. This study proposes a Particle Swarm Optimization (PSO)-optimized two-level stacking ensemble for numerical data imputation. The framework employs PSO for base-learner selection and hyperparameter optimization; the selected base learners are Random Forest and XGBoost. A two-metalearner architecture is then used in which Linear Regression captures linear dependencies and Random Forest models nonlinear interactions to generate the final imputed values. The proposed framework was evaluated on the Breast Cancer Wisconsin and Wine Quality datasets under Missing Completely at Random (MCAR) mechanisms at 30%, 20%, and 10% missingness. Performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), and processing time. Experimental results demonstrate that the proposed model consistently outperformed the standalone Random Forest and XGBoost models across all missingness levels. On the Breast Cancer dataset, the proposed model achieved RMSE values of 0.0747, 0.0627, and 0.0529 with corresponding R^2 values of 67.66%, 70.51%, and 74.63% at 30%, 20%, and 10% MCAR, respectively. Similarly, on the Wine Quality dataset, it recorded RMSE values of 0.1873, 0.1751, and 0.1681 with corresponding R^2 values of 49.54%, 52.15%, and 54.64%. Furthermore, the proposed approach outperformed a recently reported hybrid imputation method, achieving substantially lower prediction errors while maintaining acceptable computational cost. These findings demonstrate that integrating PSO optimization with a two-level stacking ensemble provides an accurate, robust, and scalable framework for numerical missing value imputation across datasets with varying degrees of missingness.

Keywords— Missing data imputation, stacking ensemble, particle swarm optimization, linear regression, random forest, hybrid machine learning.

I. INTRODUCTION

High-quality data is essential for reliable machine learning and statistical modeling. However, realworld datasets often suffer from missing values due to errors in data collection, transmission, or entry, which can reduce model accuracy and lead to biased results (Little & Rubin, 2019). Traditional approaches, such as mean imputation, deletion, or regression-based imputation, often fail to capture complex relationships and may distort the data distribution (Little, 1992; van Buuren, 2018).

Machine learning approaches such as k-nearest neighbors (KNN), decision trees, and random forests have improved imputation by modeling nonlinear relationships among features (Hastie et al., 2009; Chen & Guestrin, 2016). Empirical research has shown that machine learning approaches often provide more accurate results than traditional statistical methods. These models are better at capturing complex patterns in data. As a result, they are widely considered more effective for numerical data imputation tasks than traditional methods (Joel et al., 2025).

Ensemble learning methods, including bagging and boosting, further enhance prediction performance by combining multiple learners (Azad et al., 2025; Kotan & Kırışoğlu, 2025). Stacking ensembles, in particular, utilize heterogeneous base learners and meta-learners to learn optimal combinations of predictions (Wolpert, 1992).

Hyperparameter tuning is a critical factor for model performance. Particle Swarm Optimization (PSO) is a population-based metaheuristic that efficiently searches the hyperparameter space using the collective intelligence of a particle swarm (Kennedy & Eberhart, 1995). Previous studies have shown that hybrid frameworks combining ensemble learning and optimization algorithms outperform individual

models for predictive tasks and data imputation (Florea & Andonie, 2019).

Despite these advances, most imputation frameworks rely on single-layer stacking or do not explicitly separate linear and nonlinear relationships in the meta-learner stage. To address this gap, this study proposes a PSO-optimized two-level stacking ensemble. PSO first selects and tunes the base learners (Random Forest and XGBoost). Linear Regression (LR) then serves as the first meta-learner to capture linear patterns, and Random Forest (RF) serves as the second meta-learner to capture nonlinear interactions.

The main contributions of the study are:

- Development of a PSO-optimized stacking ensemble framework for numerical data imputation.
- Integration of a two-meta-learner architecture that explicitly captures linear and nonlinear patterns.
- Comprehensive evaluation using the Breast Cancer Wisconsin and Wine Quality datasets under varying missingness.
- Demonstration of significant improvements over existing approaches.

II. RELATED WORKS

Handling missing data has been extensively studied in statistics and machine learning. Early techniques such as mean, median, or regression imputation are simple but prone to bias (Little, 1992; van Buuren, 2018). Multiple Imputation by Chained Equations (MICE) improves on simple imputation by iteratively modeling missing variables, but its performance is sensitive to highdimensionality and complex nonlinear patterns (van Buuren, 2018).

Machine learning-based imputation methods leverage feature relationships to improve accuracy. KNN uses similarity among instances to estimate missing values (Hastie et al., 2009), while treebased models like random forests capture nonlinear interactions (Chen & Guestrin, 2016). Deep learning approaches, such as autoencoders, have also been proposed to model complex structures for imputation (Nordhausen, 2013), and transformer-based architectures have recently been used for tabular data imputation.

Ensemble learning combines multiple learners to enhance accuracy and robustness. Bagging methods, e.g., random forests, reduce variance, and boosting methods, e.g., XGBoost, focus on correcting errors sequentially (Azad et al., 2025; Kotan & Kırışoğlu, 2025). Stacking ensembles go further by using meta-learners to optimally combine predictions from heterogeneous base learners (Wolpert, 1992; Cortez et al., 2009). Research findings indicate that hybrid imputation techniques often produce better results than individual methods, particularly when dealing with complex datasets. These approaches combine the strengths of multiple models to improve the accuracy of missing data estimation. As a result, they are increasingly used in data preprocessing and analysis tasks. However, several studies primarily focus on predictive performance. Consequently, they do not always provide a detailed evaluation of computational efficiency (Amala Deepa & Lucia Agnes Beena, 2025).

Optimization algorithms, particularly PSO, have been applied to hyperparameter tuning and model selection for ensemble learning, improving generalization performance (Kennedy & Eberhart, 1995; Goodfellow et al., 2016). Recent hybrid approaches combining PSO and ensemble methods show enhanced robustness for missing value imputation (Florea & Andonie, 2019). However, multi-layer stacking frameworks explicitly capturing linear and nonlinear patterns remain underexplored.

III. PROPOSED METHODOLOGY

The proposed framework consists of three main components:

- Base Learners: Candidate models including Random Forest (RF), XGBoost, Linear Regression (LR), Ridge Regression, and Artificial Neural Networks (ANN) are optimized using PSO for hyperparameters and selection. After optimization, Random Forest and XGBoost are retained as the base learners.
- First-Level Meta-Learner (LR): Linear Regression combines predictions from base learners to capture linear relationships.
- Second-Level Meta-Learner (RF): Random Forest combines first-level predictions to capture nonlinear dependencies and improve overall accuracy.

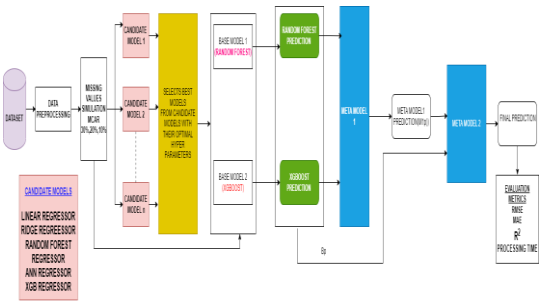

Fig. 1: Proposed Model Framework

Two-level architecture: PSO-selected RF + XGBoost base learners → Linear Regression meta-learner 1 → Random Forest meta-learner 2 → final imputed values

2. Particle Swarm Optimization (PSO)

PSO is used for base learner selection and hyperparameter tuning. Each particle represents a potential configuration (set of hyperparameters and base models).

The velocity update and position update rules are given in Equation (1) and Equation (2) respectively:

Velocity Update Equation

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 \big(pbest_i - x_i(t)\big) + c_2 r_2 \big(gbest - x_i(t)\big) \quad (1)$$

Position Update Equation

$$x_i(t+1) = x_i(t) + v_i(t+1)) \quad (2)$$

Where

- $v_i(t)$= velocity of particle iat iteration t
- $x_i(t)$ = position of particle iat iteration t
- ω = inertia weight controlling the balance between global exploration (searching new areas) and local exploitation (refining current solutions)
- c_1= cognitive coefficient, representing the particle's tendency to return to its own best-known position
- c_2= social coefficient, representing the influence of the swamp's best-known position
- $r_1 r_2$= random numbers uniformly distributed in the range [0,1], introducing stochasticity into the search process
- $pbest_i$= best position found by particle i so far (personal best)
- $gbest$ = best position found by the entire swarm (global best)

In the experiments the following PSO parameter values were used: swarm size = 20, maximum iterations = 30, inertia weight w = 0.7, cognitive coefficient c_1 = 1.5, social coefficient c_2 = 1.5. These settings provided a practical trade-off between exploration and computational cost for the candidate hyperparameter spaces of the base learners.

3. Model Formulation

The proposed PSO-Optimized Stacking Ensemble for Missing Value Imputation integrates ensemble learning with evolutionary optimization to enhance imputation performance. The framework consists of four major steps. First, the best models are selected from a pool of candidate models through Particle Swarm Optimization (PSO). Second, PSO is further applied to optimize the hyperparameters of the selected models to improve their predictive capabilities. Third, a Stacking-based ensemble is constructed by combining the optimized models as base models.

Let the selected and optimized base models be BM_1 (Random Forest) and BM_2 (XGBoost). For a given incomplete feature vector x the base predictions are obtained as:

$$\text{Bmi}\ (X) \xrightarrow{generate} \text{Bp}\ \{ f_1(X), f_2(X), \ldots f_K(X)\} \quad (3)$$

$$\text{Bp} = f(X)$$

The first-level meta-learner (Linear Regression) produces the intermediate prediction:

$$yM1 = g(f_1(X), f_2(X), \ldots f_K(X)) \quad (4)$$

Where

- X is the main dataset
- Bp is the base models predictions
- $yM1(M1p)$ is the meta model 1 prediction

The second-level meta-learner (Random Forest) receives both the base predictions and the first-level meta prediction and yields the final imputed value:

$$yM2 = g(Bp, M_1 p) \quad (5)$$

Where

- Bp is the base models predictions
- M_1p is the meat model 1 predictions
- $YM2(M2p)$is the meta model 2 prediction.

Algorithm 1: PSO-Optimized Two-Level Stacking Imputation

Input: Dataset D with simulated MCAR missing values, PSO parameters (swarm size, iterations, w, c_1, c_2)

- For each candidate base model: run PSO to obtain optimal hyperparameters and fitness (RMSE).
- Select the two best models (Random Forest and XGBoost) according to lowest RMSE.
- Train the selected base models on the observed portions of D and generate base predictions B .p
- Train Linear Regression meta-learner on B → obtain M_p 1 .p
- Train Random Forest meta-learner on the concatenated features [B , M1] → obtain $\text{final}_{p\ p}$ prediction ŷ.
- Replace missing entries in D with ŷ and return the imputed dataset.

Output: Fully imputed dataset D_imputed.

4. Evaluation Metrics

The framework is evaluated using:

- RMSE: Measures root mean squared differences between actual and imputed values.
- MAE: Measures mean absolute error between predicted and true values.
- R^2: Coefficient of determination, representing explained variance (reported as percentage).
- Processing Time (s): Wall-clock time of the complete imputation pipeline (PSO optimisation + base training + two-level meta training + prediction) measured on a single run under the Google Colab free-tier environment (Intel Xeon 2.20 GHz, 16 GB RAM).

5. Datasets

- Breast Cancer Wisconsin Dataset: Contains 569 instances with 30 features. The dataset originated from the UCI machine learning repository. Missing values were introduced artificially at 10%, 20%, and 30% MCAR.
- Wine Quality Dataset: White wine dataset with 4,898 instances and 11 physicochemical features. The dataset is from the UCI machine learning repository. Missing values were simulated similarly.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

1. Hyperparameters Optimization

After extensive evaluation, Random Forest and XGBoost models demonstrated the most consistent and superior performance across datasets.

The optimal Hyperparameters obtained via PSO are presented in Table 4.1:

Table 1 The optimized Hyperparameters obtained via PSO

Model	Optimal Hyperparameters
Random Forest (RF)	n_estimators = 169, max_depth = 15, min_samples_leaf = 1
XGBoost (XGB)	max_depth = 9, learning_rate = 0.028, n_estimators = 145, subsample = 0.867, colsample_bytree = 0.700

The optimized models were hybridized using a two-meta learners stacking ensemble strategy. This hybridization approach leverages the strengths of two base learners, the generalization capacity of Random Forest and the boosting efficiency of XGBoost models.

2. Models Imputation Performance

Breast Cancer Dataset

The models performance at different missing rates on Breast cancer Dataset is presented in Table 4.2 – 4.4 and Fig 4.1 – 4.2 respectively.

Table 2: Models Performance on Breast cancer Dataset at 30% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.0758	0.0497	65.00	0.5274
XGBOOST	0.1008	0.0683	50.82	0.1162
Proposed Stacking Model	0.0747	0.0491	67.66	3.6210

Table 3: Models Performance on Breast cancer Dataset at 20% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.0640	0.0480	70.31	0.6857
XGBOOST	0.0842	0.0566	56.90	0.1431
Proposed Stacking Model	0.0627	0.0403	70.51	3.2469

Table 4: Models Performance on Breast cancer Dataset at 10% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.0539	0.0351	73.96	0.5577
XGBOOST	0.0743	0.0492	59.31	0.1229
Proposed Stacking Model	0.0529	0.0350	74.63	3.2331

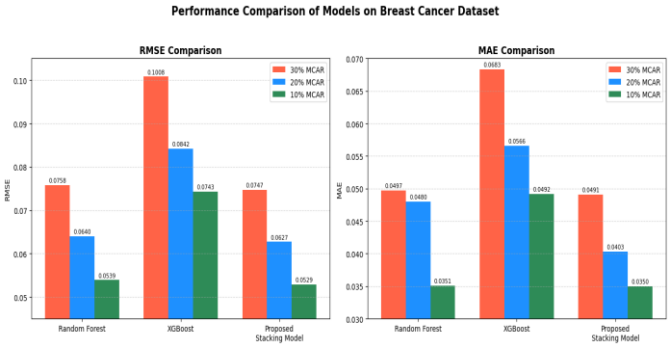

Fig 4.1 Models RMSE and MAE across the different Missing scenarios on Breast cancer Dataset

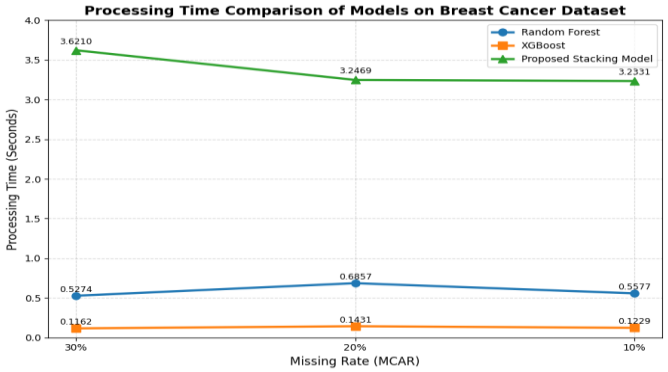

Fig 4.2 Models Processsing time across the different Missing scenarios on Breast cancer Dataset

The results in Tables 4.2–4.4 and Figures 4.1–4.2 show that the proposed Stacking model consistently outperformed the standalone Random Forest and XGBoost models across all MCAR levels on the Breast Cancer dataset. It achieved the lowest RMSE and MAE together with the highest R^2 at 30%, 20%, and 10% MCAR, indicating superior numerical imputation accuracy. Performance improved as the missing rate decreased, evidenced by declining RMSE and MAE values and increasing R^2 values. Random Forest consistently ranked second, while XGBoost produced the highest prediction errors and lowest R^2. Although the Stacking model required a longer processing time than the individual models, the additional computational cost was justified by its significantly better predictive performance. Overall, the findings demonstrate that the proposed Stacking model is the most accurate and reliable approach for numerical data imputation on the Breast Cancer dataset.

Wine Quality Dataset

The models performance on wine quality dataset at different missing rates is presented in Table 4.5–4.7 and Fig 4.3–4.4 respectively.

Table 5: Models Performance on wine quality Dataset at 30% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.1901	0.1480	48.01	1.190
XGBOOST	0.1921	0.1497	46.93	1.169
Proposed Stacking Model	0.1873	0.1429	49.54	3.012

Table 6: Models Performance on wine quality Dataset at 20% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.1812	0.1243	49.52	1.081
XGBOOST	0.1835	0.1301	47.43	1.007
Proposed Stacking Model	0.1751	0.1141	52.15	2.401

Table 7: Models Performance on wine quality Dataset at 10% MCAR

Model	RMSE	MAE	R^2 (%)	Processing Time (s)
Random Forest	0.1708	0.1201	50.95	1.000
XGBOOST	0.1803	0.1290	47.43	0.990
Proposed Stacking Model	0.1681	0.1122	54.64	1.891

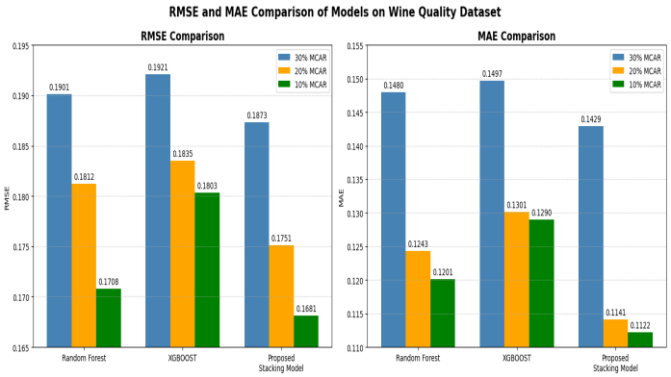

Fig 4.3 Models RMSE and MAE across the different Missing scenarios on wine quality Dataset

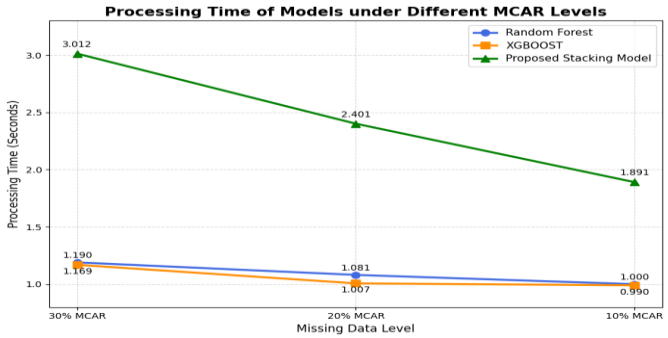

Fig 4.4 Models Processing time(s) across the different Missing scenarios on wine quality Dataset

The results in Tables 4.5–4.7 and Figures 4.3–4.4 indicate that the proposed Stacking model consistently outperformed Random Forest and XGBoost across all MCAR levels on the Wine Quality dataset. At 30%, 20%, and 10% MCAR, it achieved the lowest RMSE values of 0.1873, 0.1751, and 0.1681, with corresponding MAE values of 0.1429, 0.1141, and 0.1122, respectively. The model also recorded the highest R^2 values of 49.54%, 52.15%, and 54.64%, demonstrating superior predictive accuracy. As the missing rate decreased from 30% to 10%, RMSE and MAE declined while R^2 increased, indicating improved imputation performance. Although the Stacking model required slightly longer processing times (3.012 s, 2.401 s, and 1.891 s) than the individual models, the computational overhead was justified by the significant improvement in prediction accuracy. Overall, the findings confirm that the proposed Stacking model provides the most accurate and reliable numerical data imputation for the Wine Quality dataset.

3. Comparison with Existing Methods

To further evaluate the proposed model, the performance of the proposed stacking model was compared with an existing hybrid technique for handling missing values.

Table 8 Comparison of Proposed Model with existing work on Wine quality Dataset at 30% MCAR

Models	Type	RMSE	MAE	R^2 (%)	Processing Time (s)
Deepa and Beena (2025)	Hybrid (PoBa)	0.439	0.193	—	—
Proposed Hybrid Model	Improved Stacking (RF+XGB)	0.1873	0.1429	49.54	3.012

Table 4.8 shows that the proposed Improved Stacking Model (RF+XGB) recorded a lower RMSE of 0.1873 compared with 0.439 and a lower MAE of 0.1429 compared with 0.193 relative to the hybrid PoBa method of Deepa and Beena (2025) on the Wine Quality dataset at 30% MCAR. It should be noted that absolute RMSE values can be influenced by differences in data scaling or preprocessing; therefore the comparison is presented as indicative rather than definitive. Nevertheless, under the experimental conditions of the present study the proposed model produced substantially smaller prediction errors while also reporting an R^2 of 49.54% and a processing time of 3.012 s.

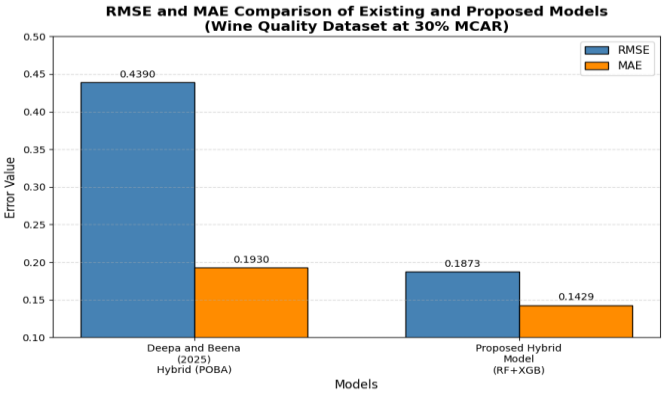

Fig 4.5 Comparison of proposed model and existing work on wine quality dataset

Fig 4.5 illustrates the reduction in RMSE and MAE achieved by the proposed stacking model relative to the PoBa hybrid approach of Deepa and Beena (2025) on the Wine Quality dataset at

30% MCAR.

V. DISCUSSION

The experimental results demonstrate that the proposed PSO-optimized two-level Stacking Model provides an effective and reliable approach for numerical missing value imputation. By combining PSO-based hyperparameter optimization with stacking ensemble learning, the framework first selects and tunes Random Forest and XGBoost as base learners, then uses Linear Regression to capture linear dependencies and Random Forest to model nonlinear interactions, ensuring robust performance across datasets. The proposed framework consistently achieved lower RMSE and MAE values and higher R^2 values than the standalone Random Forest and XGBoost models. This indicates that optimizing and integrating complementary machine learning algorithms enhances the model's ability to learn complex data patterns and produce more accurate imputations.

Across both the Breast Cancer and Wine Quality datasets, the proposed model consistently outperformed the individual models under all MCAR levels. As the percentage of missing values decreased from 30% to 10%, prediction errors steadily reduced while R^2 increased, indicating improved imputation performance with more complete data. Although the stacking framework required slightly longer processing time due to PSO optimization and the ensemble architecture, the increase in computational cost was relatively small and justified by the significant improvement in prediction accuracy. The model also

demonstrated robustness by maintaining superior performance on datasets with different sizes and complexities.

Comparison with previous studies further confirms the effectiveness of the proposed framework. The model achieved substantially lower RMSE and MAE values than the hybrid PoBa approach of Deepa and Beena (2025) on the Wine Quality dataset under the experimental conditions reported here. These improvements demonstrate that integrating PSO optimization with a stacking ensemble of Random Forest and XGBoost provides a more accurate and robust imputation strategy than the existing approach evaluated in this study.

Overall, the findings validate the proposed PSO-optimized two-level Stacking Model as an accurate, robust, and scalable framework for numerical data imputation. The model consistently maintained low prediction errors, high explanatory power, and stable performance across varying missing data levels while outperforming both standalone machine learning models and the compared hybrid imputation technique. These results highlight the potential of combining evolutionary optimization with ensemble learning to improve the accuracy and reliability of numerical missing value imputation.

VI. CONCLUSION

This study presents a PSO-optimized two-level stacking ensemble for numerical data imputation, combining base learners with Linear Regression and Random Forest meta-learners to capture linear and nonlinear relationships. Experiments on Breast Cancer and Wine Quality datasets show substantial improvements in RMSE, MAE, and R^2 compared with existing hybrid imputation methods.

The proposed framework is robust, scalable, and effective for real-world imputation tasks, particularly in scenarios where both linear and nonlinear dependencies are present in the data.

Future Works

The following directions are suggested to further enhance and extend the proposed imputation framework:

- Explore lightweight optimization and parallelization strategies to reduce computational cost.
- Extend the framework to include additional base learners.
- Investigate adaptation to other missingness mechanisms such as Missing at Random (MAR) and Missing Not at Random (MNAR).

- Authors contribution

- Conceptualization, B.I.M. and S.A.; methodology, B.I.M., S.A., Z.S.I. and A.A.A.; software, B.I.M., S.A.; validation, S.A., A.A.A. and Z.S.I.; formal analysis, B.I.M., S.A.; investigation, S.A., B.I.M. and A.A.A.; resources, B.I.M. and S.A.; data curation, A.A.A.; writing original draft preparation, B.I.M.; writing, review and editing, B.I.M., S.A., Z.S.I., and A.A.A.; visualization, B.I.M.; supervision, S.A., Z.S.I. and A.A.A.; project administration, B.I.M., S.A. and Z.S.I.; funding acquisition, B.I.M. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The breast cancer dataset used in this study is publicly available at https://doi.org/10.24432/C5DW2B and the wine quality dataset is publicly available at https://doi.org/10.1016/j.dss.2009.05.016. Both datasets can also be accessed through the UCI Machine Learning Repository.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

1. Amala Deepa V., & T. Lucia Agnes Beena. (2025). Enhancing Data Imputation in Complex Datasets Using Lagrange Polynomial Interpolation and Hot-Deck Fusion. The Scientific Temper, 16(02), 3727–3735. https://doi.org/10.58414/scientifictemper.2025.16.2.05
2. Azad, F., Bosnić, Z., & Kukar, M. (2025). Meta-Imputation Balanced (MIB): An Ensemble Approach for Handling Missing Data in Biomedical Machine Learning. 2025 IEEE 25th International Conference on Bioinformatics and Bioengineering (BIBE), 71–75. https://doi.org/10.1109/bibe66822.2025.00020
3. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
4. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785– 794. https://doi.org/10.1145/2939672.2939785
5. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling Wine Preferences by Data Mining from Physicochemical Properties. Decision Support

Systems, 47(4), 547– 553. https://doi.org/10.1016/j.dss.2009.05.016
6. Florea, A.-C., & Andonie, R. (2019). Weighted Random Search for Hyperparameter Optimization. International Journal of Computers Communications & Control, 14(2), 154–169. https://doi.org/10.15837/ijccc.2019.2.3514
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. http://www.deeplearningbook.org
8. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer New York. https://doi.org/10.1007/978-0-387-84858-7
9. Joel, L. O., Doorsamy, W., & Paul, B. S. (2025). A Comparative Study of Imputation Techniques for Missing Values in Healthcare Diagnostic Datasets. International Journal of Data Science and Analytics. https://doi.org/10.1007/s41060-025-00825-9
10. Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942–1948. https://doi.org/10.1109/icnn.1995.488968
11. Kotan, K., & Kırışoğlu, S. (2025). Cyclical Hybrid Imputation Technique for Missing Values in Data Sets. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-90964-7
12. Little, R. J. A. (1992). Regression With Missing X's: A review. Journal of the American Statistical Association, 87(420), 1227. https://doi.org/10.2307/2290664
13. Little, R. J. A., & Rubin, D. B. (2019). Statistical Analysis with Missing Data (3rd ed.). John Wiley & Sons, Inc.
14. Van Buuren, S. (2018). Flexible Imputation of Missing Data, Second Edition. Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259
15. Wolpert, D. H. (1992). Stacked Generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/s0893-6080(05)80023-1