International Journal of Advances in Intelligent Informatics
Vol. 11, No. 4, November 2025, pp. 653-666
ISSN 2442-6571

PSO-enhanced ensemble techniques for pandemic prediction and feature importance analysis

Syafrial Fachri Pane [a,1,*], Mahmud Dwi Sulistiyo [a,2], Alfian Akbar Gozali [a,3], Adiwijaya [a,4]

[a] School of Computing, Telkom University, Bandung, Indonesia
[1] fachrie@student.telkomuniversity.ac.id; [2] mahmuddwis@telkomuniversity.ac.id; [3] alfian@tass.telkomuniversity.ac.id;
[4] adiwijaya@telkomuniversity.ac.id
* corresponding author

ARTICLE INFO

Article history
Received May 7, 2025
Revised October 7, 2025
Accepted October 8, 2025
Available online November 30, 2025

Keywords
Ensemble technique
Stacking regression
Tree SHAP
PSO
Feature Importance

ABSTRACT

During the pandemic crisis that hit after 2020, Indonesia, like many other countries, faced tremendous challenges across health, the economy, and mobility. An in-depth understanding of the dynamics and changes in these areas is essential to address the impacts of the pandemic. This research is an attempt to deeply analyze the impact of the pandemic and the most effective forecasting methods based on data and phenomena. Indonesia, with its growing economy and constantly adapting health system, faces conventional economic impacts, while its health system response tries to keep up with urgent needs driven by the spread of the virus. In the context of mobility, changes in how people move and interact significantly affect virus transmission. Modeling a pandemic event with all its complexities is not easy. Even more so, in finding the right prediction method, ensemble techniques such as stacking and regression voting are emerging as promising approaches. However, deep learning and particle swarm optimization (PSO) techniques offer new innovations. The results of this study show that the ensemble vote provides the best performance in predicting confirmed positive cases and mortality based on factors of health, economic, and population mobility in Indonesia. Through feature importance analysis using MDI and Tree SHAP, we conclude that factors such as active cases, the number of vaccinations, and economic indicators, such as close IDR and close IHSG, have a significant influence on the growth of confirmed positive cases. Meanwhile, recovery factors and vaccination numbers play an important role in the growth of the number of death cases. This study confirms that a multivariate approach that considers health, economy, and mobility is the key to understanding and responding more effectively to the pandemic in Indonesia.

1. Introduction

Since the start of the pandemic in 2020, its impact has led to significant changes in health, the economy, and population mobility [1]. According to Gioachin et al., the health infrastructure has become vulnerable, the global economy has regressed, and social mobility has changed drastically [2]. Liu et al. noted that while various approaches have been used to predict pandemic impacts, many have proven ineffective in handling complex data and variable interrelationships [3].Thus, a new, reliable approach is essential to overcome data complexity.

The global pandemic poses unprecedented challenges that require innovative solutions, particularly in understanding its impacts and predicting the evolution of key characteristics. Literature highlights

https://doi.org/10.26555/ijain.v11i4.2091

various perspectives: Aktar et al. stressed the economic consequences with shifts in global consumption and production [4], while Lee et al. demonstrated the strong effect of mobility on virus transmission [5]. Ensemble methods have gained attention as promising predictive techniques [6], with Ajagbe et al. showing the strength of deep learning for large datasets [7] Optimization approaches such as PSO and interpretability frameworks like Shapley values have also been advocated, reinforcing the need for integrating diverse methodologies [8].

Given the challenge of predicting the outcome of a pandemic, the reliability of methods is critical. Ensemble techniques, such as stacking and regression voting, offer innovative solutions to overcome the complexity of the data [9]. Unlike traditional approaches, stacking combines data from multiple sources, while regression voting provides a comprehensive picture by averaging predictions. Previous studies confirm their efficacy: GBR has been applied in Germany, Japan, South Korea, and Ukraine [6], XGBR has shown strong performance in Toronto [10], RFR combined with ARIMA improved predictive accuracy [11], DTR and LR were effective in India [12], KNN and AdaBoost supported patient survival prediction [13], and SVR succeeded in forecasting cases in SAARC countries [14]. In the Indonesian context, Pane et al. analyzed the effectiveness of PPKM using multiple linear regression and dynamic factor analysis, identifying five key temporal factors (IDR exchange rate, positive cases, retail activity, transit station mobility, and recovery) with 84% accuracy and no heteroscedasticity issues, proving the effectiveness of data-driven policy [15]. When applied, ensemble techniques improve accuracy and prediction, which can help governments make effective pandemic policies [16]. Evaluating performance across R-Square, RMSE, MSE, MAE, and MAPE allows us to determine the top-performing model for forecasting confirmed positives and mortality given specific predictors.

Pandemic forecasting requires an innovative approach. According to Andrew et al., hyperparameter tuning can further optimize ensemble techniques. An effective tool for this is Particle Swarm Optimization (PSO) [17]. Combining PSO and ensemble methods gives us an accurate and highly optimized approach [18]. However, innovation continues beyond that. The global pandemic forces us to understand the impact of change, especially in the economy and mobility. Jaung et al. found that although the pandemic brought many changes, not all were caused by the virus. Therefore, we need to analyze using a technique called feature importance to help understand the model better, such as what features or variables most influence the model's predictions. The technique can be performed using MDI and Tree SHAP [19].

Understanding the feature is essential for developing reliable predictive models, not just for identifying correlational relationships [20]. In previous research, combined with PSO optimization, the ensemble method approach can improve model accuracy by identifying features [21]. For example, if population mobility is shown to have a strong causal relationship with virus spread, then this feature should be prioritized in model training. In addition, Papana et al. emphasize the importance of distinguishing between correlational relationships and relationships with true causality to improve the accuracy of predictions [22]. By focusing on feature importance and optimizing it with PSO, this research has created a model that can analyze pandemic data more accurately. This modelling allows us to implement more appropriate policies and interventions with our resources.

To our knowledge, ensemble stacking and voting methods are globally recognized advanced forecasting models. However, their application in discerning causal relationships remains underexplored, particularly in DKI Jakarta, Indonesia. This study addresses this research gap by applying optimized ensemble models to Jakarta data, combined with interpretable feature analysis, which has not been sufficiently explored in prior works. Addressing this research gap, our study seeks to predict and elucidate the underlying mechanisms driving feature importance. The key contributions of this research are as follows:

- This study explicitly aims to develop PSO-optimized ensemble models to improve pandemic prediction accuracy and to analyze feature importance for policy guidance. Development of an ensemble technique integrating stacking, voting methods, and PSO optimization is conducted to enhance the prediction accuracy of positive cases and mortality rates during the pandemic.

- In-depth exploration of feature importance through MDI and Tree SHAP techniques, uncovering subtle insights into factors influencing pandemic outcomes.
- Comparative evaluation of ensemble stacking and voting methods using statistical performance metrics such as R-Square, RMSE, MSE, MAE, and MAPE.

The paper proceeds as follows. In Section 2, we outline the research design and methods; in Section 3, we present the findings; and in Section 4, we offer concluding remarks and highlight the key contributions.

2. Method

Fig. 1 shows the steps of this research. First, we collected a dataset of COVID-19 cases, which includes health, population mobility, and economic factors. The second step is preprocessing, which involves identifying and handling missing data. Specifically, linear interpolation is applied for missing health-related data, while forward filling is used for economic data. After preprocessing, the dataset is divided using K-fold validation for training and testing purposes. A time pattern is then determined using a rolling time window technique with parameters n, d, l, s, and w. The rolling window parameters are explicitly defined as follows [23]: n represents the length of the input window in days, d is the forward prediction horizon, l denotes the number of lag variables used, s is the step size for shifting the window, and w refers to the window weight applied in historical data weighting. This explicit definition is intended to enhance transparency and facilitate model replication.

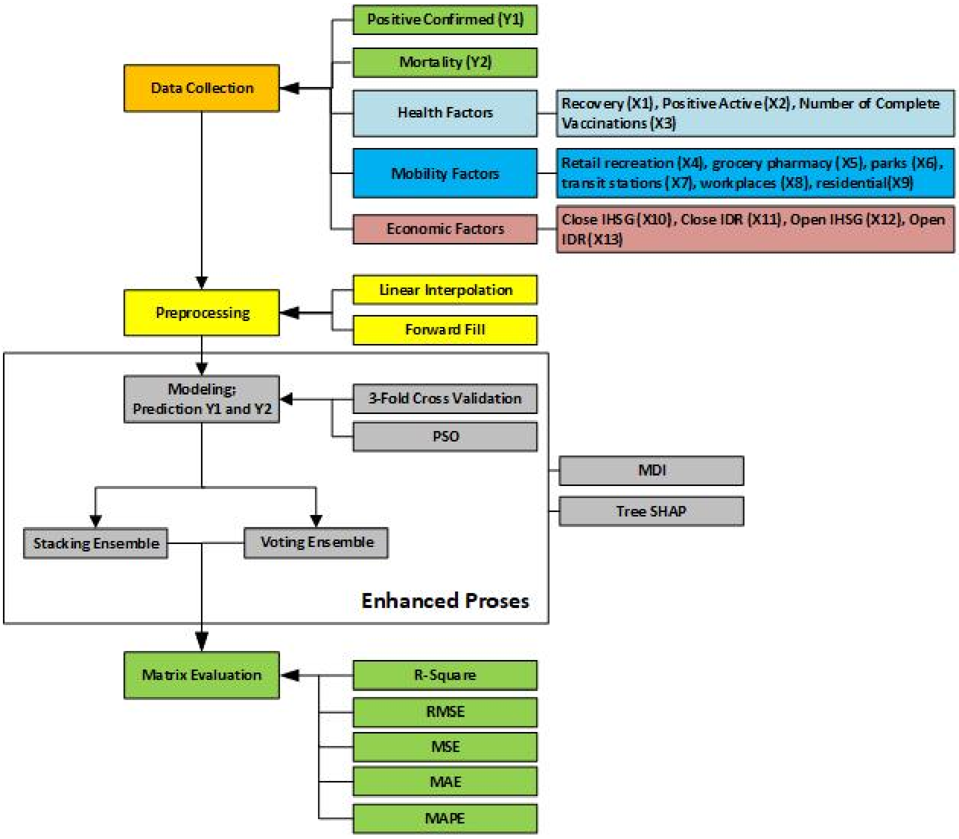

Fig. 1. The proposed research methodology.

The third step involves model development using Particle Swarm Optimization (PSO) to optimize ensemble methods, specifically ensemble stacking and voting techniques. Each model is trained to predict positive confirmed cases and mortality rates using the available health, mobility, and economic data. Finally, model performance is evaluated using multiple statistical metrics: R-Square, RMSE, MSE, MAE, and MAPE.

2.1. Data Collection

The dataset used in this study has a daily resolution, where each record corresponds to a single calendar day. In total, it comprises 960 daily observations, covering the period from March 1, 2020, to October 15, 2022, representing the most critical phases of the COVID-19 pandemic in Jakarta. The variables integrated into the dataset are grouped into three main categories: (i) health indicators (mortality, positivity rate, recovery, active cases, and vaccination numbers), (ii) population mobility indicators obtained from Google Mobility Reports (retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas), and (iii) economic indicators retrieved from Yahoo Finance (daily open and close values of the Jakarta Composite Index [IHSG] and the Indonesian Rupiah [IDR] exchange rate). Health-related data were sourced from the COVID-19 Task Force of DKI Jakarta, mobility indicators from Google Mobility Reports, and economic indicators from Yahoo Finance.

The final count of 960 records results from integrating these three heterogeneous data sources, with only dates with complete and valid entries across all variables retained after the merge. During integration, some dates were automatically excluded due to missing or inconsistent data in one or more sources (e.g., non-trading days in financial data or missing reports in mobility data). To handle minor missing values and preserve temporal continuity, two complementary imputation techniques were applied. Linear interpolation was used for continuous variables, such as health and mobility indicators, since these variables typically exhibit gradual day-to-day changes, and interpolating between adjacent points yields smooth and realistic estimates [24]. Meanwhile, forward fill (Last Observation Carried Forward, LOCF) was applied to economic indicators, reflecting the persistence of financial values during non-trading days where market activity pauses but the last known observation remains valid [25]. The focus on DKI Jakarta was made strategically, as the province serves as Indonesia's capital and the central hub of national data collection, functioning as a reference point for nationwide decision-making during the pandemic. Consequently, insights derived from Jakarta's dataset are not only statistically robust but also policy-relevant, offering valuable guidance for public health and economic interventions at the national level.

2.2. Preprocessing

During this stage, a procedure is executed to identify any missing data values. If missing values are detected, a technique known as linear interpolation is used to estimate and fill in the missing data. In the study conducted by Siregar et al.[26], interpolation is described as a method used to determine the value of a linear equation function using the principle of proportionality. The following formula may represent the mathematical expression for the interpolation.

$$l = \frac{r - t_{\text{value}}}{r - d.f} x(d.f - \text{lowest}.\, d.f). \tag{1}$$

Equation (1), it can be elucidated that the variable l represents the interpolation value, while denotes the range of t_{value} including the two nearest values d. In addition, f is involved in this context.

2.3. Modeling

Ensemble Stacking Regression is a technique that combines predictions from multiple regression models to get more accurate results [27]. Suppose we have several friends who give us their house price predictions, and then we calculate the average of their estimates to get the final prediction. In this case, each friend represents a regression model, and averaging their predictions is similar to what we do in Ensemble Stacking Regression. Thus, instead of relying on a single model, we leverage multiple models to achieve better predictions.

Suppose that we have m basic regression models. For an input χ, each model i gives a prediction $\widehat{y_j}(\chi)$. Stacking tries to find the optimal weights for each prediction of the base model so that a linear combination gives the optimal final prediction.

So, for input χ, the final prediction $\widehat{Y}(\chi)$ by the stacking method is:

$$\hat{Y}(\chi) = \alpha_1\, \hat{y}_1(\chi) + \alpha_2\, \hat{y}_2(\chi) + \cdots + \alpha_m\, \hat{y}_m(\chi) \tag{2}$$

where $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ are the weights assigned to each model. In the simplest stacking approach, the weights may be an average of the base model predictions, in which case:

$$\alpha_1 = \alpha_2 = \cdots = \alpha_m = \frac{1}{m} \tag{3}$$

Equation (2) and (3) explain that (n) represents the total number of basic regression models used in the stacking technique. (χ). represents the input data that will be predicted. $\hat{y}_j(\chi)$ represents the prediction of the i model based on the input (χ). (α_1) represents the weight for the prediction of the i model. $\hat{Y}(\chi)$ represents the final prediction of the stacking method for the input (χ). In a simple approach, all models have the same weight, which is equal to $(\frac{1}{m})$.

Ensemble Voting Regression is a technique that combines predictions from multiple regression models by taking the mean or median of the model predictions [28]. Unlike classification ensembles, where models vote on a particular class, Ensemble Voting Regression relies on each model to provide a numerical prediction. This approach reduces individual model variability and errors, resulting in more stable and accurate predictions. The mean or median of all predictions is used for ensemble voting. This technique can improve the performance of regression models and obtain more reliable predictions. Here is the formula to perform the ensemble voting:

The most common approach to Voting Regression is to take the average of predictions across all models. Given (n) regression models with each model giving a prediction (p_i) for an input (X), the final prediction $(P(X))$ of the voting regression method is given by:

$$P(X) = \frac{1}{n} \sum_{i=1} p_{i(X)} \tag{4}$$

Equation (4) explains that $(P(X))$ is the final prediction of the voting regression method for input (X). (n) is the total number of regression models in the voting technique. $(p_{i(X)})$ is the prediction of the i regression model for input (X).

2.4. Hyperparameter Tuning with PSO (Particle Swarm Optimization)

PSO is an optimization algorithm inspired by the behavior of flocks of birds or fish. It uses particles representing potential solutions to search for the optimal solution in the solution space [29], [30]. In a stacking ensemble, PSO is used to find the best weights in each regression model so that the combined predictions can be as accurate as possible before being sent to the next model. In the ensemble, PSO optimizes the weights for each regression model when combining the predictions, resulting in a weighted average that gives the best results. Using PSO, this research can optimize how ensemble stacking and voting work to produce more accurate predictions.

2.5. Feature Importance

MDI (Mean Decrease in Impurity) and Tree SHAP (Shapley Additive Explanations) are methods used to interpret machine learning models and analyze feature importance [31]. Although MDI is commonly used to calculate variable importance in decision tree models and other tree-based models [32], Tree SHAP is a more sophisticated and accurate method [23]. MDI determines importance based on the average decrease in impurity produced by the variables when dividing the tree, giving excessive preference to variables with more categories. In contrast, Tree SHAP calculates the Shapley value using game theory, which fairly distributes the contribution of each variable to the prediction for a given observation. Consider all possible combinations of variables, providing a more comprehensive interpretation, particularly when variables are correlated or have high interaction effects. Ultimately, Tree SHAP offers a more in-depth and fair analysis of feature importance in tree-based models.

2.6. Matrix Evaluation

The approach utilized in the final step of this study is based on typical statistical assessments [33], i.e., R-Square, RMSE, MAE, and MAPE [23]. According to Pane et al., the RMSE and MAPE assessments were used to measure the effectiveness of the model in predicting COVID-19 instances [34]. RMSE is used to assess model performance, whereas MAE is used to assess model prediction abilities.

3. Results and Discussion

3.1. Results

First, we do pre-processing, which uses linear interpolation techniques to fill incomplete raw data for health factors, then use forward fill for economic factors. In previous studies, interpolation was often applied to handle missing values in time series data [35]. In this work, we addressed missing values using two approaches tailored to the characteristics of the data. Linear interpolation was applied to health-related time series, as it assumes smooth variation between adjacent points and is effective for short gaps in epidemiological data [36]. For economic indicators, we used the forward fill (LOCF) method, which replaces missing entries with the most recent valid observation, reflecting the persistence of financial values until new updates are available [37]. Then, dropping na removes incomplete data in the population mobility factor. Next, the collected data are time series, so a rolling time window is used to assess the quality of the dataset by specifying the parameters. The third stage is modeling using ensemble stacking and voting techniques using regression-based machine learning models, including GBR, XGBR, RFR, DTR, KNN, SVR, AdaBoost, and LR. Then, based on matrix evaluation model regression, to improve the precision of pandemic prediction, a new ensemble technique has been developed using particle swarm optimization (PSO) as a hyperparameter tuning method with triple cross-validation.

The K-Fold results for the ensemble stacking model are shown in Fig. 2, and the results for the voting model are shown in Fig. 3. Fig. 2 shows the cross-validation results and displays the R-Square score for predicting confirmed positive cases and mortality using K=9, based on the time window parameter. The model shows stable R-Square scores above 0.85 for positive confirmed cases, except for a drop at fold 8. Mortality prediction remains consistently high (≈1.0) across all folds. In comparison, Fig. 3 for the ensemble voting model shows different results with K=6. In this study, the choice of folds in cross-validation was deliberately differentiated between stacking and voting. For stacking, we used K=9 to ensure higher variability in the training–validation splits, providing the meta-learner with stronger out-of-fold prediction signals. In contrast, for voting, we applied K=6 since this method only aggregates predictions without a meta-learner, thus not requiring highly granular partitioning and offering greater computational efficiency.

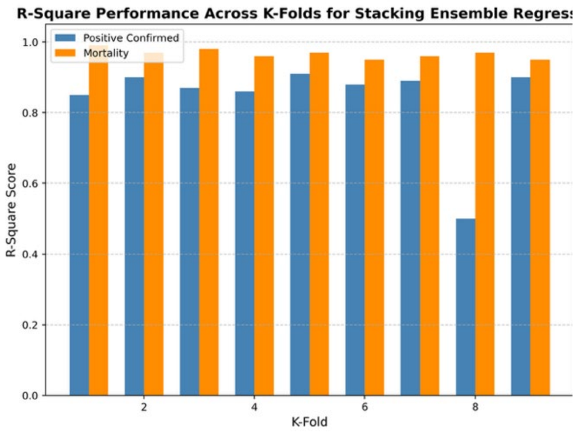

Fig. 2. R-Square on Each K-Fold for Stacking Ensemble Regression

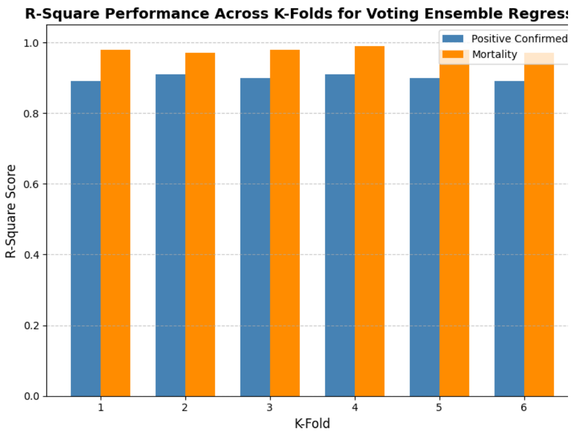

Fig. 3. R-Square on Each K-Fold for Voting Ensemble Regression

Ensemble techniques are used to predict positive confirmed cases and mortality, respectively. Fig. 4–7 present Actual vs. Predicted comparisons for two targets (Y1 and Y2) using two ensemble approaches. In Fig. 4 (Y1–Stacking) and Fig. 5 (Y2–Stacking), test-set predictions (red hollow circles) cluster closely around the ideal line $y = x$, yielding $R^2 = 0.9507$and $R^2 = 0.9952$, respectively. In contrast, Fig. 6 (Y1–Voting) and Fig. 7 (Y2–Voting) show even tighter alignment with the ideal line, with $R^2 = 0.9755$for Y1 and $R^2 = 0.9996 for Y2$. These results indicate that while both methods are accurate, Ensemble Voting is more precise and stable, particularly for Y2, which is nearly perfect, whereas Y1 still exhibits minor deviations at extreme values. Consequently, Voting is recommended for operational deployment, with continued monitoring of high-range Y1 cases. The voting ensemble achieves stable R-Square scores (0.88–0.91) for positive confirmed cases and higher, consistent values (0.96–0.99) for mortality, outperforming stacking in stability. The practice of varying K in ensemble techniques has been widely adopted in recent studies, [38]–[41] underscoring the importance of balancing methodological robustness and efficiency when selecting cross-validation schemes.

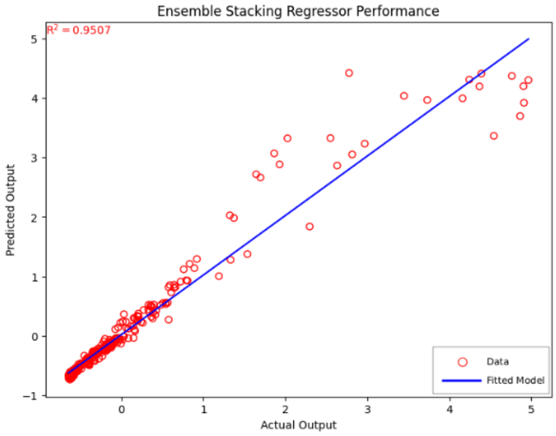

Fig. 4. Actual vs. Predicted for Y1 using the Ensemble Stacking Regressor

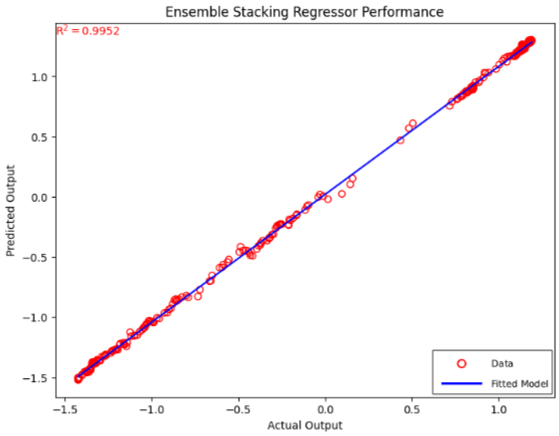

Fig. 5. Actual vs. Predicted for Y2 using the Ensemble Stacking Regressor

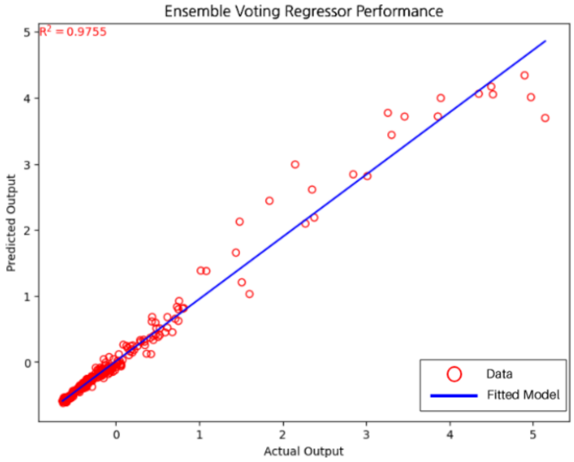

Fig. 6. Actual vs. Predicted for Y1 using the Ensemble Voting Regressor

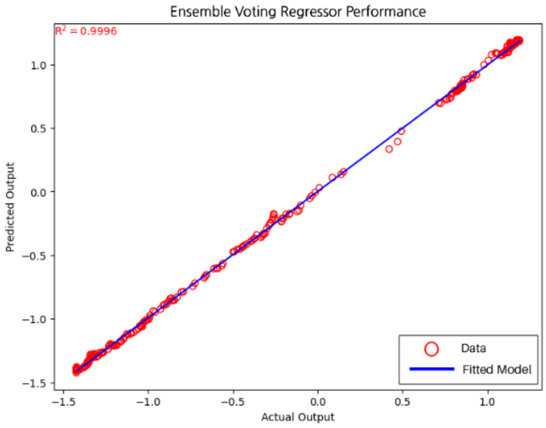

Fig. 7. Actual vs. Predicted for Y2 using the Ensemble Voting Regressor

Based on Algorithm 1, our proposed technique involves running each model in the ensemble for 10 PSO iterations, thereby enabling the identification of the optimal hyperparameter combination. PSO was employed to optimize five key parameters: the number of trees in the forest (n), the maximum depth of the tree (d), the minimum data required to split internal nodes (s), the minimum data to form leaf nodes (l), and the size of the rolling time window (w) [23]. The search ranges were defined as n:[5,50], d:[5,50], s:[2,50], l:[1,20], and w:{3,5,7}. The optimization process used a global-best (star) topology, in which particles iteratively update their positions to maximize the model's predictive performance.

Algorithm 1: PSO-Enhanced Ensemble Prediction and Feature Importance Analysis

```
Data: covid_dataset (health, economic, and mobility factors)
Result: predictions, feature_importance
covid_data ← collect_data();
Extract factors from covid_data:;
    health_data ← extract_health_data(covid_data);
    mobility_data ← extract_mobility_data(covid_data);
    economic_data ← extract_economic_data(covid_data);
missing_data ← identify_missing_values(covid_data);
if missing_data in health_data then
    health_data ← linear_interpolation(health_data);
end
if missing_data in economic_data then
    economic_data ← forward_fill(economic_data);
end
train_data, test_data ← split_data(covid_data);
time_windows ← setup_rolling_window(train_data, parameters={n, d, l, s, w});
k_fold_data ← k_fold_validation(time_windows);
models ← [GBR, XGBR, RFR, DTR, KNN, SVR, AdaBoost, LR];
stacking_model ← initialize_stacking(models);
voting_model ← initialize_voting(models);
pso ← initialize_pso();
for each model m in models do
    optimal_params ← pso.optimize(m.parameters);
    m ← update_model_with_params(m, optimal_params);
end
stacking_model.fit(train_data);
voting_model.fit(train_data);
for each ensemble model e in [stacking_model, voting_model] do
    evaluation_results ← evaluate(e, test_data, metrics=[R-Square, RMSE, MAE, MSE, MAPE]);
end
mdi_importance ← calculate_mdi(stacking_model);
tree_shap_importance ← calculate_tree_shap(stacking_model);
feature_importance ← aggregate_importance(mdi_importance, tree_shap_importance);
predictions ← predict(voting_model, test_data);
return predictions, feature_importance;
```

The fitness function was based on predictive accuracy, with scores constrained to remain within a valid range. The algorithm proceeded for a fixed number of iterations, and the best-performing parameter combinations were selected. The optimal parameters identified through this procedure are reported in Table 1, while the corresponding model performance is presented in Table 2.

Table 1. Model Regression with optimal parameter

Model	Prediction	Hyperparameter Tuning: PSO Iteration=10				
		n	*d*	*s*	*l*	*w*
Ensemble Stacking	Confirmed Positive (Y1)	14	31	25	4	7
	Mortality (Y2)	16	32	26	4	6
Ensemble Voting	Confirmed Positive (Y1)	44	19	15	5	4
	Mortality (Y2)	43	19	15	5	5

However, in this research, achieving superior excellence is not enough; an analysis of characteristic interests is needed to determine the causes of the increase in confirmed positive cases and mortality. We use the results of a good ensemble technique, namely voting. Voting is used to obtain correlation results for each factor that influences positive confirmed cases and mortality. Then, based on the analysis of feature importance using MDI and Tree SHAP techniques, we conclude that the growth in positive confirmed cases (Y1) and mortality (Y2) is significant.

Table 2. Performance matrix evaluation model regression

Country	Model	Prediction	R-Square	RMSE	MSE	MAE	MAPE
Indonesia, DKI	Ensemble	Confirmed Positive (Y1)	0.9507	0.2388	0.0570	0.1062	0.5278
Jakarta	Stacking	Mortality (Y2)	0.9952	0.0701	0.0049	0.0631	0.0893
Indonesia, DKI	Ensemble	Confirmed Positive (Y1)	0.9755	0.1546	0.0239	0.0741	0.3025
Jakarta	Voting	Mortality (Y2)	0.9996	0.0212	0.0005	0.0158	0.0334

In Fig. 8 and Table 3, the results of MDI and Tree SHAP based on positive confirmed cases (Y1) there are influencing factors such as positive cases, number of vaccinations, recovery, close IDR and close IHSG, while based on Fig. 9 and Table 4 What influences the growth in the number of mortality (Y2) are recovered cases and the number of vaccinations.

Table 3. GBR: Significant feature importance comparison using MDI and Tree SHAP for Y1

Feature	MDI	Feature	Tree SHAP
X2	Positive Active	X2	Positive Active
X3	Number of Vaccinations	X9	Close IHSG
X1	Recovery	X5	Grocery Pharmacy
X10	Close IDR	X7	Transit Stations
X9	Close IHSG	X4	Retail Recreation

Table 4. GBR: Significant feature importance comparison using MDI and Tree SHAP for Y2

Feature	MDI	Feature	Tree SHAP
X1	Recovery	X1	Recovery
X3	Number of Vaccinations	X3	Number of Vaccinations

Furthermore, the Tree SHAP analysis results identify the most important features, as shown in red and blue. Red indicates a high feature value, and blue indicates a low value. If the red dot is located on the positive side of the plot, it indicates that a high value of the feature tends to increase the prediction, and if the blue dot is located on the negative side, then vice versa. So, Fig. 10 and Table 3 show the results of the features that strongly influence the prediction of confirmed positive cases, close IDR, close IHSG, grocery pharmacy, transit station, and retail recreation. In contrast, Fig. 11 and Table 4 show that two features strongly influence the prediction of mortality cases: the recovery variable and the number of vaccinations. The last stage of the model we built is evaluated for best performance using R-Square, RMSE, MSE, MAE, and MAPE. The results of the regression model evaluation based on the matrix are shown in Table 1, and the optimal parameters are shown in Table 2.

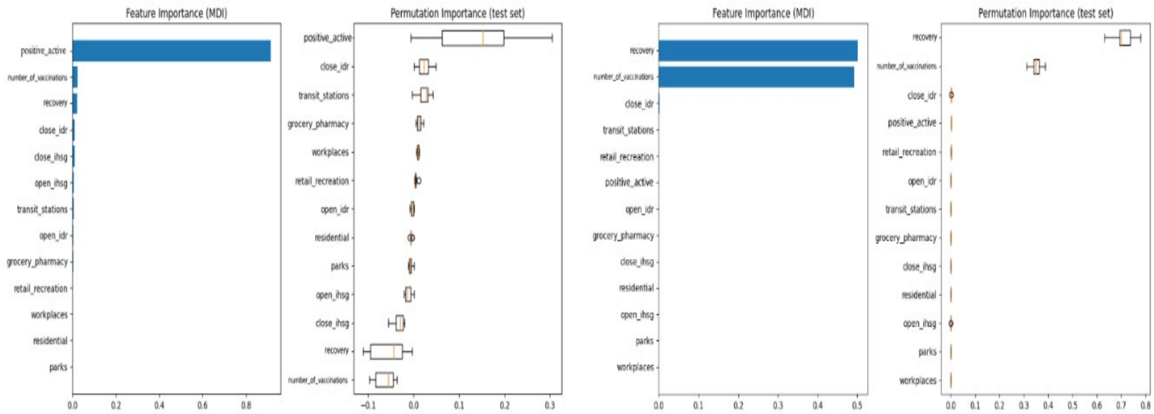

Fig. 8. MDI Permutation with GBR Model for $Y1$

Fig. 9. MDI Permutation with GBR Model for $Y2$

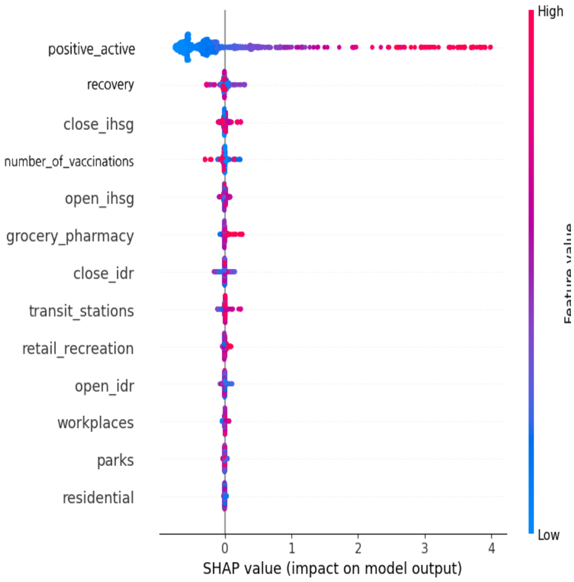

Fig. 10. Tree SHAP with GBR model for $Y1$

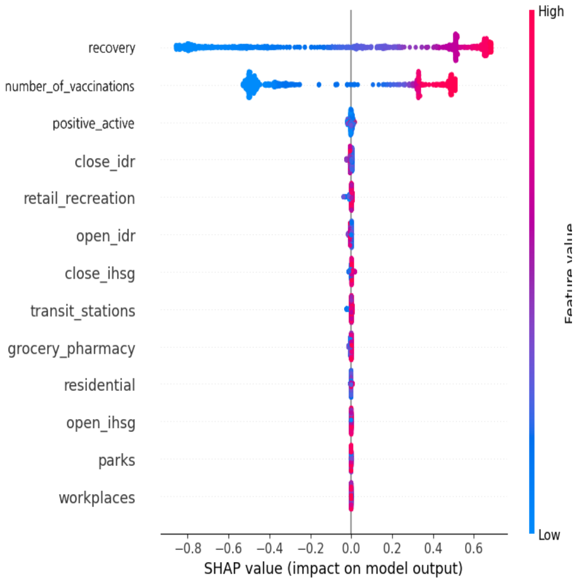

Fig. 11. Tree SHAP with GBR model for $Y2$

As shown in Table 2, the ensemble voting model consistently outperformed stacking, with higher R-Square values (0.9755 and 0.9996 compared to 0.9507 and 0.9952) and lower error metrics across RMSE, MSE, MAE, and MAPE. To ensure that these differences are not merely due to random variation, we applied statistical significance tests on the K-fold cross-validation results. Both the paired t-test and the Wilcoxon signed-rank test confirmed that the improvements of voting over stacking are statistically significant ($p < 0.05$) for all key evaluation metrics. This finding strengthens the conclusion that ensemble voting provides not only better average performance but also robust and reliable improvements over stacking, in line with recent ensemble learning studies [41], [42].

3.2. Discussion

The model developed in this study is based on multivariate data from Jakarta, Indonesia, representing health, economic, and mobility indicators. Consequently, the predictive results are highly influenced by the specific conditions of Jakarta. Nevertheless, the methodological framework employed, namely the integration of PSO-optimized ensemble techniques with feature importance analysis, is generic and can potentially be applied to other regions with similar data characteristics. This opens the door to future studies evaluating the model's transferability across different regional contexts. Although this study is limited to Jakarta data, this limitation is mitigated by the robust methodologies employed. We applied K-Fold cross-validation to evaluate model performance by partitioning the dataset into multiple subsets, allowing repeated training and validation cycles that minimize overfitting and provide a more generalized performance assessment [23], [43]. In addition, we implemented a rolling time-window evaluation, which trains models on a fixed-size window of historical data and tests them on subsequent observations. This method is particularly effective for time-series forecasting, as it simulates real-world scenarios in which models must continuously adapt to new data [23].

As shown in Table 5, while traditional models such as ARIMA are easy to interpret, LSTM excels at capturing long-term dependencies, yet both remain limited in handling nonlinear and multivariate dynamics. Similarly, XGBoost and hybrid approaches demonstrate strong performance but require complex tuning and provide limited interpretability. In contrast, our proposed PSO-Enhanced Ensemble integrates multiple regressors, optimizes hyperparameters using PSO, and incorporates feature importance analysis (Tree SHAP and MDI). This results in superior accuracy (R^2 up to 0.9996) and greater policy relevance, offering both predictive performance and interpretability for decision-making.

Based on the results of our research, we can predict modeling using ensemble techniques, namely stacking and voting, optimized with PSO based on the evaluation matrix. The ensemble voting technique outperforms the stacking technique across R-Square, RMSE, MSE, MAE, and MAPE. Then, the results of the GBR model-based feature importance analysis indicate that variables associated with positive confirmed cases and mortality are significant, as seen in Table 3 and Table 4.

Table 5. Comparative summary of Pandemic models

Ref	Dataset	Time Rolling Window	PSO	Hybrid	Modelling	Accuracy
[12]	India	×	×	×	ARIMA	80 % to 85 %
[14]	SAARC countries	×	×	×	LSTM	MAE 0.12 %
[34]	Indonesia, Jakarta dan Jawa Barat	×	×	×	LSTM dan ARIMA	91% to 94%
[10]	Toronto	×	×	×	XGboost	94%
[11]	Indian	×	×	√	RFR + ARIMA	0.95 %
Propose Method	Indonesia, Jakarta	√	√	√	Ensemble Voting	97 % to 99 %

The text discusses the need for feature-interest analysis to understand the factors contributing to the increase in confirmed positive cases and mortality. The researchers used the GBR regression model to determine the correlation between factors, positive cases, and mortality. MDI and Tree SHAP techniques identified the influencing factors for positive cases and mortality growth. The results showed that positive cases, the number of vaccinations, recovery, closed IDR, and closed IHSG influenced the increase in positive cases. The recovered cases and the number of vaccinations were significant factors in the increase in mortality. The Tree SHAP analysis highlighted the most important features using red and blue, with red indicating a high value and blue a low one. The features with the highest predictive power are presented in Fig.10 and Table 3.

The results of the feature importance analysis using MDI and Tree SHAP techniques indicate differences in their assessments of the importance of features related to positive confirmed cases. At the same time, they give consistent assessments for features related to mortality cases. This disparity arises because Tree SHAP accounts for feature interactions, whereas MDI does not. Strong feature interactions can explain this discrepancy in feature importance. Additionally, MDI can be biased towards features with more categories, meaning that it may consider both features important in cases of high correlation, whereas Tree SHAP may prioritize only one. If there is a significant difference in the number of categories between features, this could also lead to discrepancies in the feature importance results. In conclusion, if speed and lower accuracy are acceptable, MDI may suffice. However, for a more accurate and comprehensive analysis, where computation time and memory are not a concern, Tree SHAP is the preferred choice.

The contribution of this research is demonstrated not only by achieving high R^2 values but also by the insights gained from interpreting key variables that influence pandemic dynamics. The feature importance analysis enables the identification of dominant factors, such as workplace mobility, vaccination rates, and financial market volatility, which can serve as a basis for public policy formulation. Thus, this study emphasizes not only predictive accuracy but also the practical utility of the model's outcomes in supporting decision-making.

4. Conclusion

The study argues that a deep understanding of changes in health, economy, and mobility during a pandemic is essential. The study also reviewed the forecasting methods used, the impact of the pandemic on the economy and the health system, and the spread of the virus. The results show that ensemble methods such as stacking and regression voting are promising techniques, with deep learning and particle swarm optimization (PSO) providing additional innovation. Ensemble voting can accurately predict confirmed cases and COVID-19 mortality based on health, economic, and population mobility factors. The regression models were optimized using PSO, yielding different optimal parameter sets for Ensemble Stacking and Ensemble Voting. Both models performed well in predicting confirmed cases (Y1) and mortality (Y2) in DKI Jakarta. Ensemble Stacking achieved high accuracy with R-square values of 0.9507 for confirmed cases and 0.9952 for mortality. Ensemble Voting performed even better, reaching 0.9755 and 0.9996 for the exact predictions. Error values such as RMSE, MSE, and MAE were lower in Ensemble Voting, indicating stronger overall performance. Overall, Ensemble Voting is the most accurate model, particularly for mortality prediction. Furthermore, analysis using MDI and Tree SHAP techniques shows that positive cases affect the growth of positive confirmed cases, the number of vaccinations, recovered cases, closed IDR, closed IHSG, grocery pharmacy, transit station, and retail recreation. Meanwhile, factors such as recovery and vaccination rates affect the growth in mortality cases.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] H. Alizadeh, A. Sharifi, S. Damanbagh, H. Nazarnia, and M. Nazarnia, “Impacts of the COVID-19 pandemic on the social sphere and lessons for crisis management: a literature review,” *Nat. Hazards 2023 1173*, vol. 117, no. 3, pp. 2139–2164, Apr. 2023, doi: 10.1007/S11069-023-05959-2.

[2] F. Gioachin, I. Marx, and S. Scherer, “Stratification of poverty risk: The importance of social class in four European countries,” *Soc. Sci. Res.*, vol. 110, no. February, p. 102814, Feb. 2023, doi: 10.1016/J.SSRESEARCH.2022.102814.

[3] Y. Liu, S. Srivastava, Z. Huang, and F. J. Vázquez-Abad, “Pandemic model with data-driven phase detection, a study using COVID-19 data,” *J. Oper. Res. Soc.*, vol. 74, no. 2, pp. 450–464, 2023, doi: 10.1080/01605682.2021.1982652;JOURNAL:JOURNAL:TJOR20.

[4] M. A. Aktar, M. M. Alam, and A. Q. Al-Amin, “Global economic crisis, energy use, CO2 emissions, and policy roadmap amid COVID-19,” *Sustain. Prod. Consum.*, vol. 26, no. April, pp. 770–781, Apr. 2021, doi: 10.1016/J.SPC.2020.12.029.

[5] K. S. Lee and J. K. Eom, “Systematic literature review on impacts of COVID-19 pandemic and corresponding measures on mobility,” *Transp. 2023 515*, vol. 51, no. 5, pp. 1907–1961, Apr. 2023, doi: 10.1007/S11116-023-10392-2.

[6] D. Chumachenko, I. Meniailov, K. Bazilevych, T. Chumachenko, and S. Yakovlev, “Investigation of Statistical Machine Learning Models for COVID-19 Epidemic Process Simulation: Random Forest, K-Nearest Neighbors, Gradient Boosting,” *Comput. 2022, Vol. 10, Page 86*, vol. 10, no. 6, p. 86, May 2022, doi: 10.3390/COMPUTATION10060086.

[7] S. A. Ajagbe and M. O. Adigun, “Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review,” *Multimed. Tools Appl. 2023 832*, vol. 83, no. 2, pp. 5893–5927, May 2023, doi: 10.1007/S11042-023-15805-Z.

[8] Y. Al-Smadi, M. Eshtay, A. Al-Qerem, S. Nashwan, O. Ouda, and A. A. Abd El-Aziz, “Reliable prediction of software defects using Shapley interpretable machine learning models,” *Egypt. Informatics J.*, vol. 24, no. 3, p. 100386, Sep. 2023, doi: 10.1016/J.EIJ.2023.05.011.

[9] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, “Ensemble Learning for Disease Prediction: A Review,” *Healthc. 2023, Vol. 11, Page 1808*, vol. 11, no. 12, p. 1808, Jun. 2023, doi: 10.3390/HEALTHCARE11121808.

[10] C. Feng, G. Kephart, and E. Juarez-Colunga, "Predicting COVID-19 mortality risk in Toronto, Canada: a comparison of tree-based and regression-based machine learning methods," *BMC Med. Res. Methodol. 2021 211*, vol. 21, no. 1, pp. 267–88, Nov. 2021, doi: 10.1186/S12874-021-01441-4.

[11] A. B. Chelani and S. Gautam, "The influence of meteorological variables and lockdowns on COVID-19 cases in urban agglomerations of Indian cities," *Stoch. Environ. Res. Risk Assess. 2022 369*, vol. 36, no. 9, pp. 2949–2960, Jan. 2022, doi: 10.1007/S00477-021-02160-4.

[12] R. Kumari *et al.*, "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," *Big Data Min. Anal.*, vol. 4, no. 2, pp. 65–75, Jun. 2021, doi: 10.26599/BDMA.2020.9020013.

[13] R. Mazloumi, S. R. Abazari, F. Nafarieh, A. Aghsami, and F. Jolai, "Statistical analysis of blood characteristics of COVID-19 patients and their survival or death prediction using machine learning algorithms," *Neural Comput. Appl. 2022 3417*, vol. 34, no. 17, pp. 14729–14743, May 2022, doi: 10.1007/S00521-022-07325-Y.

[14] A. S. Saha, M. Haque, and M. G. R. Alam, "Covid-19 Confirmed-Cases Prediction in SAARC Countries through Machine Learning," in *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Sep. 2021, pp. 889–894, doi: 10.1109/ICIRCA51532.2021.9544964.

[15] S. F. Pane, A. Adiwijaya, M. D. Sulistiyo, and A. A. Gozali, "Multi-Temporal Factors to Analyze Indonesian Government Policies regarding Restrictions on Community Activities during COVID-19 Pandemic," *JOIV Int. J. Informatics Vis.*, vol. 7, no. 4, pp. 2263–2269, Dec. 2023, doi: 10.62527/JOIV.7.4.2415.

[16] M. Humayun and A. Alsayat, "Prediction Model for Coronavirus Pandemic Using Deep Learning," *Comput. Syst. Sci. Eng.*, vol. 40, no. 3, pp. 947–961, Sep. 2021, doi: 10.32604/CSSE.2022.019288.

[17] T. McAndrew and N. G. Reich, "Adaptively stacking ensembles for influenza forecasting," *Stat. Med.*, vol. 40, no. 30, pp. 6931–6952, Dec. 2021, doi: 10.1002/SIM.9219.

[18] A. Jafar and M. Lee, "High Accuracy COVID-19 Prediction Using Optimized Union Ensemble Feature Selection Approach," *IEEE Access*, vol. 12, pp. 122942–122958, 2024, doi: 10.1109/ACCESS.2024.3424231.

[19] W. Jaung, "Changes in human-nature relations during pandemic outbreaks: a big data analysis," *Sci. Total Environ.*, vol. 768, no. May, p. 144530, May 2021, doi: 10.1016/J.SCITOTENV.2020.144530.

[20] D. Uzun Ozsahin, E. Precious Onakpojeruo, B. Bartholomew Duwa, A. G. Usman, S. Isah Abba, and B. Uzun, "COVID-19 Prediction Using Black-Box Based Pearson Correlation Approach," *Diagnostics 2023, Vol. 13, Page 1264*, vol. 13, no. 7, p. 1264, Mar. 2023, doi: 10.3390/DIAGNOSTICS13071264.

[21] S. S. Band *et al.*, "Novel Ensemble Approach of Deep Learning Neural Network (DLNN) Model and Particle Swarm Optimization (PSO) Algorithm for Prediction of Gully Erosion Susceptibility," *Sensors 2020, Vol. 20, Page 5609*, vol. 20, no. 19, p. 5609, Sep. 2020, doi: 10.3390/S20195609.

[22] A. Papana, "Connectivity Analysis for Multivariate Time Series: Correlation vs. Causality," *Entropy 2021, Vol. 23, Page 1570*, vol. 23, no. 12, p. 1570, Nov. 2021, doi: 10.3390/E23121570.

[23] Y. Pan, L. Zhang, Z. Yan, M. O. Lwin, and M. J. Skibniewski, "Discovering optimal strategies for mitigating COVID-19 spread using machine learning: Experience from Asia," *Sustain. Cities Soc.*, vol. 75, no. December, p. 103254, Dec. 2021, doi: 10.1016/J.SCS.2021.103254.

[24] S. M. Ribeiro and C. L. Castro, "Missing Data in Time Series: A Review of Imputation Methods and Case Study," *Learn. Nonlinear Model.*, vol. 20, no. 1, pp. 31–46, Oct. 2022, doi: 10.21528/LNLM-VOL20-NO1-ART3.

[25] H. T. Pham *et al.*, "Handling Missing Data in COVID-19 Incidence Estimation: Secondary Data Analysis," *JMIR Public Heal. Surveill.*, vol. 10, no. 1, p. e53719, Aug. 2024, doi: 10.2196/53719.

[26] R. Ruli, A. Siregar, T. Djatna, S. Sarmose, M. P. Manullang, and I. Saputra, "Double Exponential Smoothing Berimputasi LOCF Dan Linear Interpolation Dalam Akurasi Peramalan Harga Harian Emas," *KILAT*, vol. 10, no. 1, pp. 208–222, Apr. 2021, doi: 10.33322/KILAT.V10I2.1200.

[27] M. Shahhosseini, G. Hu, and H. Pham, "Optimizing ensemble weights and hyperparameters of machine learning models for regression problems," *Mach. Learn. with Appl.*, vol. 7, no. March, p. 100251, Mar. 2022, doi: 10.1016/J.MLWA.2022.100251.

[28] S. Kumari, D. Kumar, and M. Mittal, “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,” *Int. J. Cogn. Comput. Eng.*, vol. 2, no. June, pp. 40–46, Jun. 2021, doi: 10.1016/J.IJCCE.2021.01.001.

[29] J. Tang, G. Liu, and Q. Pan, “A Review on Representative Swarm Intelligence Algorithms for Solving Optimization Problems: Applications and Trends,” *IEEE/CAA J. Autom. Sin.*, vol. 8, no. 10, pp. 1627–1643, Oct. 2021, doi: 10.1109/JAS.2021.1004129.

[30] B. Chen, L. Cao, C. Chen, Y. Chen, and Y. Yue, “A comprehensive survey on the chicken swarm optimization algorithm and its applications: state-of-the-art and research challenges,” *Artif. Intell. Rev.*, vol. 57, no. 7, p. 170, Jul. 2024, doi: 10.1007/S10462-024-10786-3/FIGURES/14.

[31] R. B. Singh and H. Zhuang, “Measurements, Analysis, Classification, and Detection of Gunshot and Gunshot-like Sounds,” *Sensors 2022, Vol. 22, Page 9170*, vol. 22, no. 23, p. 9170, Nov. 2022, doi: 10.3390/S22239170.

[32] A. Purnomo, A. G. Putrada, R. Habibi, and Syafrianita, “MDI and PI XGBoost regression-based methods: regional best pricing prediction for logistics services,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 22, no. 5, pp. 1157–1166, May 2024, doi: 10.12928/TELKOMNIKA.V22I5.26037.

[33] S. Kumar, T. Kolekar, K. Kotecha, S. Patil, and A. Bongale, “Performance evaluation for tool wear prediction based on Bi-directional, Encoder–Decoder and Hybrid Long Short-Term Memory models,” *Int. J. Qual. Reliab. Manag.*, vol. 39, no. 7, pp. 1551–1576, Jul. 2022, doi: 10.1108/IJQRM-08-2021-0291.

[34] S. F. Pane, Adiwijaya, M. D. Sulistiyo, and A. A. Gozali, “LSTM and ARIMA for Forecasting COVID-19 Positive and Mortality Cases in DKI Jakarta and West Java,” in *2022 7th International Conference on Informatics and Computing, ICIC 2022*, 2022, pp. 1–6, doi: 10.1109/ICIC56845.2022.10006959.

[35] V. K. R. Chimmula and L. Zhang, “Time series forecasting of COVID-19 transmission in Canada using LSTM networks,” *Chaos, Solitons & Fractals*, vol. 135, p. 109864, Jun. 2020, doi: 10.1016/j.chaos.2020.109864.

[36] X. Che *et al.*, “Linearly interpolating missing values in time series helps little for land cover classification using recurrent or attention networks,” *ISPRS J. Photogramm. Remote Sens.*, vol. 212, pp. 73–95, Jun. 2024, doi: 10.1016/J.ISPRSJPRS.2024.04.021.

[37] K. Kotan and S. Kırışoğlu, “Cyclical hybrid imputation technique for missing values in data sets,” *Sci. Reports 2025 151*, vol. 15, no. 1, p. 6543, Feb. 2025, doi: 10.1038/s41598-025-90964-7.

[38] R. K. Halder, M. A. Lima, M. N. Uddin, M. A. Islam, and A. Saha, “Integrated feature selection-based stacking ensemble model using optimized hyperparameters to predict breast cancer with smart web application,” *Clin. eHealth*, vol. 8, pp. 146–161, Dec. 2025, doi: 10.1016/J.CEH.2025.08.001.

[39] Y. Huang, G. Chen, J. Gou, Z. Fan, and Y. Liao, “A hybrid feature selection and aggregation strategy-based stacking ensemble technique for network intrusion detection,” *Appl. Intell. 2024 551*, vol. 55, no. 1, p. 28, Nov. 2024, doi: 10.1007/S10489-024-06015-7.

[40] Z. J. Muhsin *et al.*, “Highly efficient stacking ensemble learning model for automated keratoconus screening,” *Eye Vis. 2025 121*, vol. 12, no. 1, p. 25, Jun. 2025, doi: 10.1186/S40662-025-00440-6.

[41] S. Q. Sultan, N. Javaid, N. Alrajeh, and M. Aslam, “Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross-Validation: A Symmetric Approach,” *Symmetry 2025, Vol. 17, Page 185*, vol. 17, no. 2, p. 185, Jan. 2025, doi: 10.3390/SYM17020185.

[42] R. K. Halder, M. A. Lima, M. N. Uddin, M. A. Islam, and A. Saha, “Integrated feature selection-based stacking ensemble model using optimized hyperparameters to predict breast cancer with smart web application,” *Clin. eHealth*, vol. 8, no. December, pp. 146–161, Dec. 2025, doi: 10.1016/J.CEH.2025.08.001.

[43] M. A. Jones, R. Faiz, Y. Qiu, and B. Zheng, “Improving mammography lesion classification by optimal fusion of handcrafted and deep transfer learning features,” *Phys. Med. Biol.*, vol. 67, no. 5, p. 054001, Feb. 2022, doi: 10.1088/1361-6560/AC5297.