Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 2, April 2026, pp. 1174∼1185
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.10549

Attention-enhanced wasserstein GAN for agricultural market data imputation

Ulima Inas Shabrina, Riyanarto Sarno, Ratih Nur Esti Anggraini, Agus Tri Haryono
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

Received Apr 23, 2025
Revised Nov 14, 2025
Accepted Dec 6, 2025

Keywords:

Agricultural market data
Attention mechanism
Crop price prediction
Deep learning
Missing data imputation
Time-series forecasting
Wasserstein generative adversarial network

ABSTRACT

Crop price prediction in ASEAN markets is hindered by incomplete and inconsistent data, making data imputation essential. This study introduces the Wasserstein generative adversarial imputation network with attention (WGAIN+Att) to improve data quality for forecasting. Four configurations—GAIN, GAIN+Att, WGAIN, and WGAIN+Att—were evaluated on rice, corn, and soybean datasets (1961–2023). Results show that WGAIN+Att, particularly when attention is applied across all matrices (x, m, and z), achieved the best imputation performance, minimizing mean absolute error (MAE) and preserving statistical distributions, with optimal results at a 0.1 missing rate and 0.9 hint rate. In predictive tasks, GAIN-based imputations combined with convolutional neural networks (CNN)-long short-term memory networks (LSTM)-gated recurrent units (GRU) models consistently outperformed others in forecasting accuracy, achieving lower MAE and root mean squared error (RMSE). The findings highlight the role of attention in stabilizing imputation and ensuring realistic reconstructions, while also showing that aligning imputation with forecasting objectives improves agricultural price predictions.

Corresponding Author:

Riyanarto Sarno
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
Email: riyanarto@its.ac.id

1. INTRODUCTION

Stable food prices are vital for economic growth, poverty reduction, and food security, particularly in the ASEAN region [1]. Agricultural commodity price prediction is crucial, especially as climate change affects food production, livelihoods, nutrition, and policymaking in East and Southeast Asia [2]. Fluctuations in rice, corn, and soybean prices can disrupt markets, making accurate predictions essential for stability and policy decisions [3]. However, datasets used for price predictions often contain missing values, compromising model accuracy [4]. Standard methods, such as discarding incomplete rows, are unsuitable for time-series data, as they reduce representativeness and increase bias. Many machine learning models struggle with missing data, and existing automated solutions lack precision, highlighting the need for effective imputation techniques to improve predictions [5].

Generative adversarial imputation networks (GAIN) provide a foundation for imputing missing data by capturing complex patterns and dependencies [6]. The Wasserstein generative adversarial imputation net-

work (WGAIN) enhances both stability and performance by leveraging the Wasserstein distance, which quantifies the minimal effort required to transform one probability distribution into another. Moreover, incorporating attention mechanisms further enhances the imputation process by focusing on the most relevant features, effectively capturing essential structural relationships within the data and selectively prioritizing influential variables [7].

By integrating the stability of WGAIN with the precision of attention mechanisms, this study aims to develop a novel imputation model to address the complex and varied missing data patterns characteristic of ASEAN agricultural datasets. This approach is expected to yield more accurate and reliable predictions for commodity prices, ultimately supporting strategic decision-making in the ASEAN agrarian sector.

2. LITERATURE REVIEW

Missing data can hinder analyses, bias results, and reduce prediction reliability [8]. This issue is particularly critical when a substantial portion of data is missing, as it complicates interpretation and model training. Missing values—caused by entry errors, unavailability, or collection issues—are common in real-world tabular datasets [9], [10]. Missing data mechanisms include missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [11]. Strategies for handling missing data include: listwise deletion, which removes incomplete rows but reduces sample size; model adjustments that account for missingness; and data imputation, which replaces missing values with estimated ones. Listwise deletion is problematic for high-dimensional datasets [12]. Imputation helps maintain balance and reduce bias using statistical estimation [13].

2.1. Imputation methods

Imputation methods are categorized as discriminative or generative as shown in Table 1. Discriminative approaches (e.g., mean, median, and random forest) directly predict missing values from observed data but may oversimplify feature relationships and introduce bias, especially in skewed distributions [14]. Generative approaches model the joint distribution of observed and missing data, using techniques such as expectation–maximization (EM) [15], multiple imputation by chained equations (MICE) [16], denoising autoencoders (DAE) [17], and generative adversarial networks (GANs) [18]. These methods capture richer dependencies and can better preserve multivariate structure; however, they typically require more computation and may face convergence or scalability challenges in high dimensions. Classical methods like EM and MICE remain popular but rely heavily on distributional assumptions, while deep learning–based methods such as DAE and GANs relax such constraints and excel in capturing nonlinear patterns. Given the need to balance flexibility, accuracy, and the ability to handle complex dependencies in high-dimensional data, we adopt a GAN-based imputation method (GAIN) [19].

Table 1. Comparison of discriminative and generative imputation methods

Method	Type	What's good	What's bad	Works well when
Mean/median [20]	Discriminative	Very fast; trivial; median robust to outliers.	Destroys variance; ignores correlations; mean fails on skew.	Low missingness, quick baseline, near-symmetric data.
Random forest (missForest) [21]	Discriminative	Captures nonlinearity; handles mixed types; few assumptions.	Slow, memory-heavy; biased on rare categories; no uncertainty.	Tabular data with learnable dependencies, moderate size.
EM [15]	Generative	Statistically sound; converges under correct model; gives parameter uncertainty.	Strong assumptions; sensitive init; local optima; poor scaling.	Gaussian-like data, modest dimension.
MICE [16], [22]	Generative (via conditionals)	Flexible model choice; captures uncertainty; widely used.	Slow in high p; tricky diagnostics; model misspecification hurts; weak for strong nonlinearity.	Mixed-type clinical/survey data, moderate p.
DAE [17], [23]	Generative	Learns nonlinear manifolds; scalable; weak assumptions.	Needs large data; tuning-heavy; risk of over-smoothing; no uncertainty.	Large datasets with nonlinear structure.
GAIN [19]	Generative (GAN)	Preserves multivariate structure; fewer assumptions; strong empirical results; uses missingness mask.	Complex tuning; unstable under MNAR; needs normalization/capacity control.	High-dimensional tabular data, MCAR/MAR, enough samples.

2.2. Generative adversarial imputation network

GAIN adapts GANs for missing data imputation by estimating the conditional distribution of missing values from observed data [19]. Its generator G imputes missing values using observed data $\widetilde{X}$, mask matrix

M, and random noise Z, producing X in (1). The completed dataset $\hat{X}$ preserves observed values and fills in missing ones with generated estimates (2). Unlike traditional GANs, the discriminator D does not classify entire samples as real or fake, but instead evaluates each entry as observed or imputed. This is achieved through a hint matrix H ((3), which partially reveals observed values to prevent trivial discrimination and guide D in making informed predictions.

$$X = G\left(\widetilde{X}, M, (1-M) \odot Z, \theta_g\right) \tag{1}$$

$$\hat{X} = M \odot \widetilde{X} + (1-M) \odot X \tag{2}$$

$$H = B \odot M + 0.5 \cdot (1-B) \tag{3}$$

The training objective balances discriminator accuracy with generator imputation quality. The adversarial loss $L(D,G)$ in (4) ensures D distinguishes between observed and imputed values, penalizing incorrect predictions in both cases. The generator's loss L_G (5) combines imputation accuracy on missing entries (L_I) with reconstruction of observed values (L_O), weighted by hyperparameter α. This formulation enables GAIN to produce realistic imputations while maintaining consistency with observed data, leveraging the hint mechanism to enhance robustness and prevent trivial solutions.

$$L(D,G) = E\left[X, M, H\right]\left[M^T \log D\left(\hat{X}, H\right) + (1-M)^T \log\left(1 - D\left(\hat{X}, H\right)\right)\right] \tag{4}$$

$$L_G = \sum_{j=1}^{k} L_I\left(M^{(j)}, \widehat{M^{(j)}}, B^{(j)}\right) + \alpha L_O\left(\widetilde{X^{(j)}}, \widehat{X^{(j)}}\right) \tag{5}$$

2.3. Wasserstein generative adversarial network for imputation

The Wasserstein GAN (WGAN) enhances the stability of GAN training by incorporating the Wasserstein distance, which captures the minimal effort needed to convert one distribution into another. This approach helps prevent mode collapse, ensures more stable learning, and produces higher-quality samples. Building on this concept, the Wasserstein GAIN (WGAIN) adapts WGAN techniques for the task of imputing missing data [24]. Like GAIN, it uses observed data $\widetilde{X}$, mask M, and noise Z, with generator G imputing missing values. The key difference is that the discriminator D estimates the Wasserstein distance and employs gradient penalties for stability. Two hint matrices are used: H_{real} for observed values and H_{fake} for imputed values, both partially masked with 0.5 to prevent trivial solutions.

WGAIN is trained with a min–max game between G and D as shown in (6). The discriminator loss, shown in (7), enforces Wasserstein distance with a gradient penalty λ, while the Generator minimizes a weighted sum of imputation loss L_I, shown in (8), and reconstruction loss L_O, demonstrated in (8), balanced by α, shown in (9). This ensures stable training and realistic imputations consistent with observed data.

$$\min_G \max_D L(D,G) = E_{H \sim P_{H_{\text{real}}}}[D(X,H)] - E_{H \sim P_{H_{\text{fake}}}}[D(X,H)] \tag{6}$$

$$L_D = L(D,G) + \lambda E_{X_I,H}\left[\left(\|\nabla_{X_I} D(X_I,H)\|_2 - 1\right)^2\right] \tag{7}$$

$$L_I = -E_{H \sim P_{H_{\text{fake}}}}[D(X,H)] \tag{8}$$

$$L_O(x,x') = \sum_{i=1}^{d} m(i) \cdot (x^{(i)} - x'^{(i)})^2 \tag{9}$$

$$L_G = L_I + \alpha L_O \tag{10}$$

2.4. Attention mechanism

The attention mechanism enables models to focus on the most relevant input components by assigning adaptive weights to different parts of the data [7]. The standard Scaled Dot-Product Attention computes similarity between queries (Q) and keys (K), normalizes via softmax, and produces a weighted sum of values (V) as in (11). This formulation is equivariant to reordering rows of Q and invariant to reordering the key-value pairs, ensuring consistent behavior across input permutations as shown in (12).

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{11}$$

$$\text{softmax}(ADB) = A\,\text{softmax}(D)\,B \tag{12}$$

Multi-head attention extends this mechanism by applying multiple attention heads in parallel, each with its own learnable weight matrices. The outputs are concatenated and linearly transformed (13), where each head is defined by (14). This design captures diverse relationships across queries, keys, and values, enhancing representation power while preserving permutation equivariance and invariance properties.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{13}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{14}$$

2.5. Prediction model

Deep learning models, such as convolutional neural networks (CNN), long short-term memory networks (LSTM), and gated recurrent units (GRU), are utilized to evaluate the quality of imputed data. CNN are effective for time-series forecasting, leveraging convolutional, pooling, and fully connected layers [25]. The convolution operation is defined in (15):

$$c_t = \tanh\left(x_t * w_t + b_t\right) \tag{15}$$

The LSTM networks address the vanishing gradient issue using forget, input, and output gates, as shown in (16) to (18) [25]:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{16}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{17}$$

$$h_t = o_t * \tanh\left(c_t\right) \tag{18}$$

The GRU simplifies LSTM with reset and update gates [26], defined in (19):

$$h_t = (1 - u_t) * h_{t-1} + u_t * \tanh\left(W_{\tilde{h}} * [f_t * h_{t-1}, x_t]\right) \tag{19}$$

2.6. Evaluation

The evaluation of the imputation model considers both error metrics and statistical data distribution [27]. Root mean squared error (RMSE), mean absolute error (MAE), and mean squared error (MSE) are used as core error metrics for imputation evaluation. RMSE emphasizes large deviations, highlighting significant errors in the imputed data, while MAE is more robust to outliers, providing a balanced view of overall error. MSE, with its smooth gradient properties, supports nuanced error analysis and is particularly useful for optimization purposes [28], [29]. Alongside these metrics, the statistical alignment of the imputed data with the original data distribution is examined to ensure that the model not only minimizes error but also preserves realistic data characteristics.

For the prediction model RMSE, MAE, and MSE are also applied as evaluation metrics. These measures assess the accuracy and reliability of predictions, focusing on how well the model captures underlying patterns and reduces deviation from observed values. While RMSE and MSE prioritize significant errors for correction, MAE offers an additional layer of evaluation by being less sensitive to extreme values, thus providing a robust metric for evaluating predictive accuracy in diverse scenarios.

3. METHOD

The study begins by collecting datasets on rice, corn, and soybean prices across ASEAN countries to ensure market coverage. After data collection, preparation is performed to organize and standardize formats, addressing inconsistencies. Preprocessing follows, including normalization, outlier handling, and encoding of categorical variables to make the data model-ready. Missing values are imputed using advanced techniques like GAIN, GAIN with Attention (GAIN+Attention), WGAIN, and WGAIN with Attention (WGAIN+Attention) to retain underlying patterns.

The predictive analysis uses a stacked CNN-LSTM-GRU model, leveraging CNN for local patterns and LSTM-GRU layers for temporal forecasting. Model performance is evaluated using the metrics RMSE, MAE, and MSE to compare results on imputed and non-imputed datasets, assessing the impact of the imputation techniques on model accuracy in predicting commodity prices. The complete process is illustrated in Figure 1.

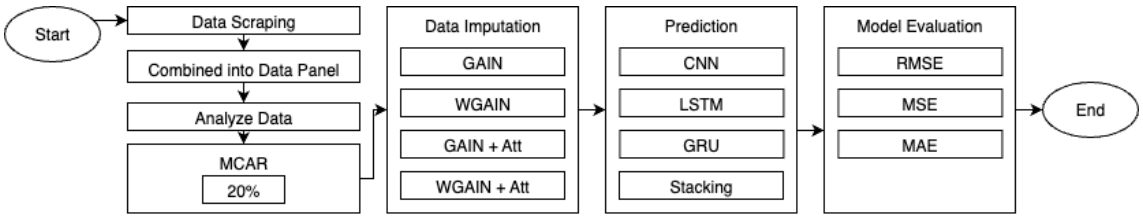


Figure 1. Method flow

3.1. Data preparation

The dataset, sourced from the FAO (1961–2023), covers three key crops—rice, soya, and corn—with 16 indicators including production, trade, population, and income variables. Organized in a panel format, it supports both cross-country and time-series analysis. Table 2 summarizes the dataset, showing 693 records per crop, average non-null counts, overall means, and variability across indicators.

Table 2. Agricultural dataset grouped by crops

Crops	Total row	Missing value (%)	Mean	Std
Corn	693	21.06	2,768,863.5	2,844,157
Rice	693	19.05	4,472,683.8	4,703,288
Soya	693	27.56	3,973,012.4	3,092,221

3.2. Imputation of missing values

This study employs four imputation methods to handle missing data: GAIN, GAIN with Attention (GAIN+Attention), WGAIN, and WGAIN with Attention (WGAIN+Attention). The results obtained in [30] mentioned that attention can improve the data imputation accuracy. The generator and discriminator used in this case are multi-layer perceptron (MLP) layers with Leaky-ReLU activation. The WGAIN+Attention algorithm was compared to GAIN, GAIN+Attention, and WGAIN. For a better evaluation of the algorithm, not only are the MSE, MAE, and RMSE of the imputation performance tested, but also the prediction performance, both showing the quality of the imputations. All experiments were performed with 20% of the data points MCAR. Table 3 shows the data imputation model parameters.

Figure 2 shows the WGAIN+Attention model. Generator: real data x, noise z, and mask m pass through a Multi-head Attention Layer, then dense layers with Leaky ReLU and Sigmoid to produce outputs. Critic: actual data x and hint matrix h are processed through dense layers to distinguish real vs. imputed data. Multi-head Attention projects inputs into query, key, and value, applies scaled dot-product attention, and merges results. To stabilize training, a gradient penalty interpolates between real and generated data, computes gradients, and regularizes the critic's loss.

3.3. Evaluation model

Imputation quality was assessed using RMSE, MSE, and MAE, where lower values indicate better accuracy. RMSE penalizes large errors more heavily, MSE measures the average squared error, and MAE captures the average absolute error, treating all deviations equally. To ensure distributional consistency, differences

in mean, standard deviation, and min–max range between actual and imputed data were analyzed, preserving central tendency, variability, and realistic value ranges [30].

Table 3. Data imputation model layer parameters

Model	Description
GAIN [19]	Generator: Input (x, z, m, h) → Masking Operation → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3, Sigmoid) → Output Discriminator: Input (x, h) → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3) → Output
GAIN+Attention [30]	Generator: Input (x, z, m, h) → Attention Layer → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3, Sigmoid) → Output Discriminator: Input (x, h) → Concatenation → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3) → Output Multi-head Attention: Input x (input data) → Linear Projections (wq, wk, wv) (create Q, K, and V using dense layers) → Split Heads → Scaled Dot Product Attention → Final Dense Layer → Output
WGAIN [24]	Generator: Input (x, z, m, h) → Masking Operation → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3, Sigmoid) → Output Critic (Discriminator): Input (x, h) → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3) → Output Gradient Penalty: Input (real, fake, mask) → Interpolation → Critic Evaluation → Gradient Calculation → Penalty Calculation → Output
WGAIN+Attention	Generator: Input (x, m, and z) (real data x noise z, mask m) → Attention Layer → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3, Sigmoid) → Output Critic (Discriminator): input (x, h) → Dense Layer (fc1, Leaky_ReLU) → Dense Layer (fc2, Leaky_ReLU) → Dense Layer (fc3) → Output Multi-head Attention: Input (x) → Linear Projection (Query, Key, Value) → Split into Heads → Scaled Dot-Product Attention → Final Linear Projection → Output Gradient Penalty: input (real data, fake data) → Interpolation → Compute Critic's Output → Gradient Calculation → Penalty Calculation → Add Penalty to Critic's Loss

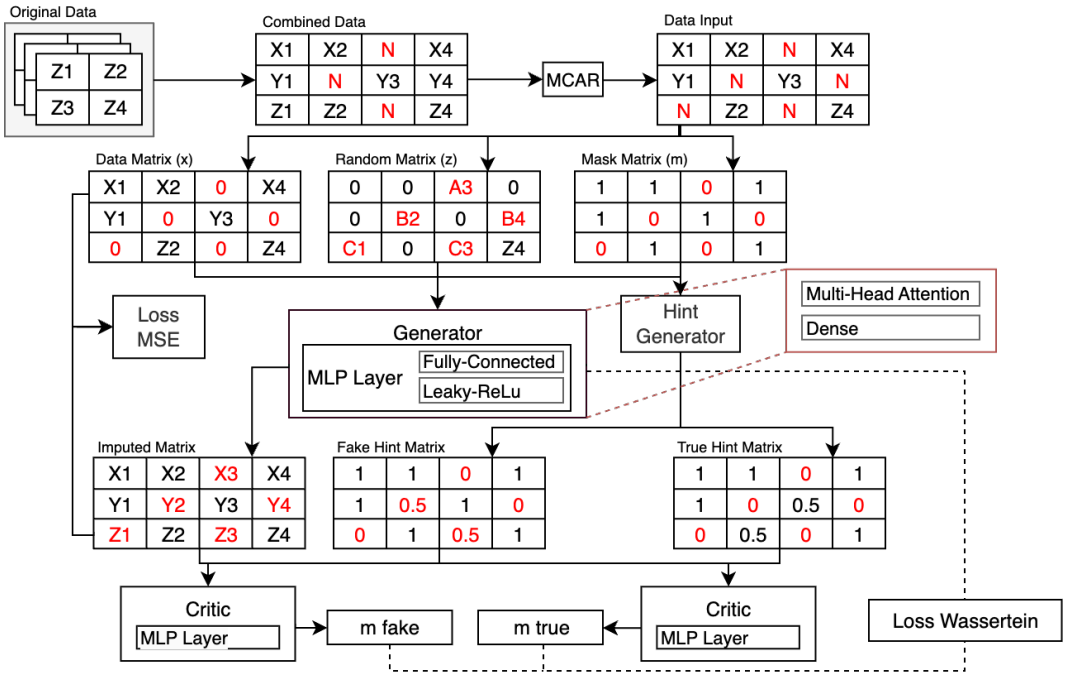


Figure 2. Wasserstein GAIN+Attention mechanism

Predictive performance was evaluated using four models: CNN, LSTM, GRU, and a stacked CNN-LSTM-GRU ensemble, all with identical configurations. CNN employed a Conv1D layer, while LSTM and GRU used their respective recurrent layers, each followed by a Dense layer for forecasting. The ensemble

combined outputs from all three models to improve accuracy. This evaluation aimed to balance predictive precision with the preservation of original data characteristics for downstream analysis.

4. RESULT AND DISCUSSION

This section evaluates the GAIN-based imputation approach on soya, rice, and corn datasets, analyzing the impact of missing rates, hint rates, and attention mechanisms to identify optimal configurations for handling missing data and improving predictive accuracy in agricultural panel data models.

4.1. Experiment setup

In these experiments, three datasets—soya, rice, and corn—are utilized, and the results are averaged across three datasets. The experiments are structured to explore different imputation approaches and examine the influence of model components on performance. This study uses a set of hyperparameters for the data imputation process, including the Leaky ReLU activation function, a learning rate of 0.0001, 10,000 iterations, a batch size of 64, the Adam optimizer, a regularization parameter (α) of 1, and a gradient penalty (λ) of 5. The experimental design evaluates imputation accuracy through three key scenarios: missing and hint rates in GAIN, the influence of the attention mechanism in GAIN, and predictive models. These scenarios aim to examine the effectiveness of GAIN and assess its influence on predictive accuracy.

4.2. Missing and hint rates in generative adversarial imputation network

The experiment was conducted across datasets for three crops, with results averaged across all. As the missing rate increased from 0.1 to 0.3, RMSE, MSE, and MAE rose, indicating higher reconstruction errors. Higher hint rates, such as 0.9, consistently improved performance, particularly at lower missing rates. For example, at a missing rate of 0.1, a hint rate of 0.9 achieved the lowest errors, while a hint rate of 0.7 led to poorer outcomes. The best overall performance was observed at a missing rate of 0.1 and a hint rate of 0.8, as shown in Table 4.

Table 4. Miss rates and hint rates experiment results

Miss rate	Hint rate	RMSE	MSE	MAE
0.1	0.9	1,374,556	3,827,248,734,791	455,114
0.1	**0.8**	**1,126,247**	**2,407,941,753,481**	**391,891**
0.1	0.7	1,178,866	2,682,880,738,436	409,519
0.2	0.9	1,525,714	4,482,774,288,633	448,395
0.2	0.8	1,553,494	4,727,397,866,979	459,571
0.2	0.7	1,613,715	5,202,041,120,164	459,629
0.3	0.9	1,602,259	4,888,101,286,762	446,366
0.3	0.8	1,724,638	5,808,766,637,399	477,656
0.3	0.7	1,723,424	5,818,647,618,608	467,043

4.3. Influence of the attention mechanism in generative adversarial imputation network

The placement of the attention mechanism within the GAIN model significantly influences imputation performance, as reflected in variations in RMSE, MSE, and MAE. This experiment evaluated different configurations of the GAIN model with a missing rate of 0.2 and a hint rate of 0.9, aligning with real-world missing data patterns (19% to 27%). Attention was applied to individual matrices—data (x), mask (m), and random (z)—and across all matrices (x, m, and z). Table 5 shows that while the baseline GAIN model performed well with an RMSE of 1,525,714 and a MAE of 448,395, the WGAIN+Attention model applied across all matrices achieved the best performance overall, with the lowest MAE of 445,376 and a competitive RMSE of 1,525,239.

Table 5. Attention mechanism in GAIN results

Model	RMSE	MSE	MAE
GAIN	1,525,714	4,482,774,288,633	448,395
GAIN+Attention	1,547,592	4,721,613,508,274	459,787
WGAIN	1,646,880	5,458,767,654,672	469,837
WGAIN+Attention (x, m, and z)	**1,525,239**	**4,516,212,569,319**	445,376
WGAIN+Attention (x)	1,690,906	5,515,409,120,802	**400,747**
WGAIN+Attention (m)	1,777,780	6,467,528,139,541	479,297
WGAIN+Attention (z)	1,725,365	5,449,050,082,831	533,968

The WGAIN+Attention (x, m, and z) configuration balanced numerical accuracy and stability, preserving the dataset's inherent variability while minimizing over-smoothing. For example, Corn's standard deviation after imputation is 2,484,432, closely aligned with the original data. In contrast, attention applied to individual matrices resulted in less stable performance, such as WGAIN+Attention (m), inflating the mean difference for Corn to 2,195,647, and WGAIN+Attention (z), increasing the standard deviation to 4,301,490. Additionally, the Wasserstein distance improved alignment with the original data distribution, as shown by the reduced standard deviation difference for Soya to 195,501.

Table 6 highlights the statistical characteristics of the dataset before and after imputation, providing a detailed analysis for each crop—Corn, Rice, and Soya. The actual data exhibited large standard deviations and wide ranges, showing its natural variability. Post-imputation, the WGAIN+Attention (x, m, and z) model effectively preserved these characteristics, reducing deviations from the original mean and standard deviation compared to other configurations. For instance, Corn's standard deviation after imputation with the value of 2,484,432 closely aligned with the original data of 2,844,157, while maintaining the lowest MAE. Similarly, the model showed a closer alignment to original data characteristics for Rice and Soya.

Table 6. Data statistics before and after imputation

Category	Type	Max	Mean	Min	Std dev	Diff. mean	Diff. std
Actual data	Corn	3,029,221	2,768,863	429	2,844,157	0	0
	Rice	10,867,858	4,472,683	1,075	4,703,288	0	0
	Soya	1,414,561	3,973,012	709	3,092,221	0	0
GAIN	Corn	4,515,679	2,855,082	4,243,298	2,437,352	86,219	406,805
	Rice	11,507,527	3,706,085	4,056,760	5,583,194	766,598	879,906
	Soya	1,638,069	4,669,933	3,671,631	4,533,052	696,921	1,440,831
GAIN+Attention	Corn	4,606,865	1,718,096	5,883,816	3,821,362	1,050,767	977,205
	Rice	10,827,908	3,223,902	1,776,238	5,616,085	1,248,780	912,797
	Soya	1,711,279	4,067,302	2,571,493	3,721,587	**94,290**	629,366
WGAIN	Corn	3,979,534	1,669,513	3,828,931	3,766,083	1,099,350	921,926
	Rice	11,207,485	4,180,169	4,539,880	5,709,706	292,514	1,006,418
	Soya	1,728,500	4,275,279	4,270,848	2,896,720	302,266	195,501
WGAIN+Attention	Corn	5,328,985	2,743,049	3,785,567	2,484,432	**25,814**	**359,725**
(x, m, and z)	Rice	11,957,686	3,436,618	3,949,744	4,946,270	1,036,065	242,982
	Soya	2,003,824	5,355,912	3,246,505	4,553,268	1,382,900	1,461,047
WGAIN+Attention	Corn	4,365,435	3,090,998	3,332,623	2,157,109	322,135	687,048
(x)	Rice	11,100,109	3,614,946	3,830,521	5,345,441	857,737	642,153
	Soya	1,515,883	2,686,596	3,232,289	4,022,415	1,286,416	930,194
WGAIN+Attention	Corn	4,642,748	573,216	4,463,479	3,604,488	2,195,647	760,331
(m)	Rice	11,013,394	4,235,607	2,650,678	4,862,477	237,076	**159,189**
	Soya	1,671,919	5,260,996	3,637,538	3,515,478	1,287,983	423,257
WGAIN+Attention	Corn	4,209,255	2,881,400	4,136,791	4,301,490	112,537	1,457,333
(z)	Rice	10,398,410	4,665,473	3,519,986	4,910,131	**192,789**	206,843
	Soya	1,670,448	4,290,917	3,269,127	3,338,896	317,904	246,675

WGAIN+Attention (x, m, and z) model stands out as the most robust imputation strategy, balancing error minimization and data integrity across multiple metrics. By leveraging attention mechanisms across all matrices, this configuration enhanced the model's ability to capture complex relationships in the data. In contrast, configurations focusing on individual matrices (x, m, or z) often led to overfitting or instability. The integration of attention mechanisms across all matrices within the WGAIN model provides a superior approach to imputation.

The WGAIN+Attention (x, m, and z) model delivers the best overall performance. For Corn, it reduces the standard deviation to 2,484,432 while maintaining mean and range values consistent with the original data trends. Additionally, this model achieves the lowest MAE of 445,376 and a competitive RMSE of 1,525,239, ensuring both numerical accuracy and realistic data distributions. By integrating attention mechanisms across multiple matrices (x, m, and z), it effectively balances error reduction and statistical integrity. In contrast, applying attention mechanisms individually to matrices (x, m, and z) often results in instability or oversimplification. For example, WGAIN+Attention (m) introduces instability in Corn with a high difference in mean (2,195,647), while WGAIN+Attention (z) inflates the standard deviation for Corn to 4,301,490. Overall, the WGAIN+Attention (x, m, and z) configuration emerges as the most robust choice, offering consistent and representative imputations suitable for downstream applications.

4.4. Predictive models

The experiment was conducted on three different crop datasets, where the performance of the predictive models was evaluated under the same conditions. The data imputation process was applied using WGAIN+Attention, and the predictive performance was assessed on imputed datasets and non-imputed datasets (where missing values were replaced with 0). The models were trained using 80% of the data and tested on the 20%, with a lookback of 10 and a horizon of 5. Each model was trained for five epochs with a batch size of 16, with the Adam optimizer. The evaluation metrics used are RMSE, MAE, and MSE, which were computed for each crop. Finally, the reported results are the average values across all three crops, providing a generalized comparison of model performance under imputed and non-imputed conditions.

Table 7 highlights the strengths of predictive models with different imputation methods. CNN performs best with baseline GAIN, achieving the lowest RMSE of 0.08360, MSE of 0.00716, and MAE of 0.05274, showing its effectiveness in capturing patterns but limited adaptability with complex imputations. LSTM with GAIN provides the best overall accuracy, showing the lowest RMSE of 0.08049 and MSE of 0.00671, while WGAIN performs better for MAE of 0.04754. GRU also scores well with GAIN, though slightly behind LSTM.

Table 7. Predictive model results

Model	Metric	Actual data	GAIN	GAIN+Att	WGAIN	WGAIN+Attention (x, m, and z)
CNN	RMSE	0.11064	**0.08360**	0.10109	0.09212	0.09193
	MSE	0.01257	**0.00716**	0.01070	0.00887	0.00874
	MAE	0.06392	**0.05274**	0.06721	0.05452	0.05581
LSTM	RMSE	0.10052	**0.08049**	0.09284	0.08225	0.08373
	MSE	0.01050	**0.00671**	0.00921	0.00691	0.00727
	MAE	**0.04834**	0.05255	0.05784	0.04754	0.04802
GRU	RMSE	0.10051	**0.08248**	0.09402	0.08809	0.08746
	MSE	0.01046	**0.00695**	0.00933	0.00785	0.00798
	MAE	**0.04910**	0.04744	0.05921	0.05281	0.05201
Stacking	RMSE	0.10604	**0.08959**	0.09587	0.09254	0.09072
	MSE	0.01160	**0.00827**	0.00967	0.00892	0.00853
	MAE	0.05862	0.05803	0.06366	**0.05803**	0.05801

The stacking ensemble shows balanced results with GAIN, showing it as a versatile but potentially overfit model when handling more complex imputations like WGAIN+Att. LSTM with GAIN is the top choice for overall accuracy. WGAIN+Attention (x, m, and z) with LSTM or Stacking excels for specific point accuracy, emphasizing the need to align model and imputation choices with specific prediction goals. This analysis highlights that while GAIN provides a strong baseline for imputation across models, the choice of model and imputation method should be aligned with specific prediction goals, such as overall error reduction or pointwise accuracy improvement.

5. CONCLUSION

Based on the experimental results, the hypothesis that combining the WGAIN model with an attention mechanism improves data imputation quality is largely supported. The study demonstrated that lower missing rates and higher hint rates consistently led to better imputation performance, with the best results observed at a missing rate of 0.1 and a hint rate of 0.9. The addition of the attention mechanism in WGAIN helped improve the stability and consistency of imputed data, particularly when attention was applied across all matrices (x, m, and z), achieving the lowest MAE and a competitive RMSE.

The placement of the attention mechanism was crucial in driving model performance. While the GAIN+Attention model showed moderate improvement, the WGAIN+Attention (x, m, and z) configuration provided the best balance of error reduction and statistical integrity. This configuration not only minimized the MAE but also ensured that the imputed data preserved realistic distributions, such as mean and standard deviation, critical for agricultural datasets. The results indicate that a more nuanced application of attention enhances the model's ability to maintain data consistency, making it especially valuable of real-world farm data where data distributions are complex.

In predictive tasks, GAIN with CNN, LSTM, GRU, or stacking models consistently outperformed other configurations, minimizing MAE and achieving competitive RMSE. WGAIN+Attention (x, m, and z) configuration proved to be the most effective in data imputation. This highlights the importance of aligning

the choice of imputation methods with overall error reduction or enhancing point-wise accuracy. Future work should focus on handling higher missing rates (e.g., 30%) and addressing challenges such as consecutive missing values to improve further the robustness and applicability of the approach in real-world scenarios.

FUNDING INFORMATION

This research is funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 4299/B3/DT.03.08/2025 & No. 3029/PKS/ITS/2025) and the National Research and Innovation Agency (BRIN) under RIIM Kompetisi Gelombang 10 Program.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Agus Tri Haryono	✓	✓		✓						✓				
Riyanarto Sarno	✓			✓	✓	✓	✓		✓			✓	✓	✓
Ratih Nur Esti Anggraeni	✓	✓	✓	✓	✓	✓	✓			✓		✓		
Ulima Inas Shabrina	✓	✓	✓	✓		✓		✓	✓	✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

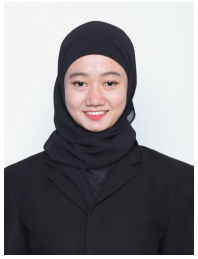
The data that support the findings of this study will be available in [https://github.com/shabrinaiu/Wasserstein-GAIN-Attention/tree/main].

REFERENCES

[1] P. Sundram, "Food security in asean: progress, challenges and future," *Frontiers in Sustainable Food Systems*, vol. 7, 2023, doi: 10.3389/fsufs.2023.1260619.
[2] H.-I. Lin, Y.-Y. Yu, F.-I. Wen, and P.-T. Liu, "Status of food security in east and southeast asia and challenges of climate change," *Climate*, vol. 10, no. 3, p. 40, 2022, doi: 10.3390/cli10030040.
[3] S. Gu and M. Li, "Forecasting corn futures prices using the lstm model," *Financial Economics Research*, vol. 1, no. 2, 2024, doi: 10.70267/st9q2c95.
[4] S. Shah, M. Telrandhe, P. Waghmode, and S. Ghane, "Imputing missing values for dataset of used cars," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, 2022, pp. 1–5, doi: 10.1109/ASIANCON55314.2022.9908600.
[5] M. Ishaq, S. Zahir, L. Iftikhar, M. F. Bulbul, S. Rho, and M. Y. Lee, "Machine learning based missing data imputation in categorical datasets," *IEEE Access*, vol. 12, pp. 88 332–88 344, 2023, doi: 10.1109/ACCESS.2024.3411817.
[6] W. Khan, N. Zaki, A. Ahmad, M. Masud, L. Ali, N. Ali, and L. Ahmed, "Mixed data imputation using generative adversarial networks," *IEEE Access*, vol. 10, pp. 124 475–124 490, 2022, doi: 10.1109/ACCESS.2022.3218067.
[7] J. Zhao, C. Rong, C. Lin, and X. Dang, "Multivariate time series data imputation using attention-based mechanism," *Neurocomputing*, vol. 542, p. 126238, 2023, doi: 10.1016/j.neucom.2023.126238.
[8] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity," *Decision Analytics Journal*, vol. 9, 2023, doi: 10.1016/j.dajour.2023.100341.
[9] M. D. Samad, S. Abrar, and N. Diawara, "Missing value estimation using clustering and deep learning within multiple imputation framework," *Knowledge-Based Systems*, vol. 249, 2022, doi: 10.1016/j.knosys.2022.108968.
[10] W.-C. Lin, C.-F. Tsai, and J. Zhong, "Deep learning for missing value imputation of continuous data and the effect of data discretization," *Knowledge-Based Systems*, vol. 239, p. 108079, 2022, doi: 10.1016/j.knosys.2021.108079.

[11] B. Gomer and K. Yuan, "A realistic evaluation of methods for handling missing data when there is a mixture of mcar, mar, and mnar mechanisms in the same dataset," *Multivariate Behavioral Research*, vol. 58, pp. 988 – 1013, 2023, doi: 10.1080/00273171.2022.2158776.
[12] C.-H. Cheng, Y.-F. Kao, and H.-P. Lin, "A financial statement fraud model based on synthesized attribute selection and a dataset with missing values and imbalanced classes," *Applied Soft Computing*, vol. 108, p. 107487, 2021, doi: doi.org/10.1016/j.asoc.2021.107487.
[13] S. C. Lotspeich and T. P. Garcia, "Extrapolation before imputation reduces bias when imputing censored covariates," *Journal of Computational and Graphical Statistics*, vol. 34, no. 4, pp. 1322–1330, 2025, doi: 10.1080/10618600.2024.2444323.
[14] H. Ou, Y. Yao, and Y. He, "Missing data imputation method combining random forest and generative adversarial imputation network," *Sensors (Basel, Switzerland)*, vol. 24, no. 4, 2024, doi: 10.3390/s24041112.
[15] D. Wang, S. Zhang, M. Gan, and J. Qiu, "A novel em identification method for hammerstein systems with missing output data," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2500–2508, 2020, doi: 10.1109/TII.2019.2931792.
[16] R. Wu, S. D. Hamshaw, L. Yang, D. W. Kincaid, R. Etheridge, and A. Ghasemkhani, "Data imputation for multivariate time series sensor data with large gaps of missing data," *IEEE Sensors Journal*, vol. 22, no. 11, pp. 10 671–10 683, 2022, doi: 10.1109/JSEN.2022.3166643.
[17] R. C. Pereira, M. S. Santos, P. P. Rodrigues, and P. H. Abreu, "Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes," *Journal of Artificial Intelligence Research*, vol. 69, pp. 1255–1285, 2020, doi: 10.1613/jair.1.12312.
[18] D. Neves, J. Alves, M. Naik, A. Proença, and F. Prasser, "From missing data imputation to data generation," *Journal of Computational Science*, vol. 61, p. 101640, 2022, doi: 10.1016/j.jocs.2022.101640.
[19] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: Missing data imputation using generative adversarial nets," in *Proceedings of the 35th International Conference on Machine Learning*. PMLR, vol. 80, 10–15 Jul 2018, pp. 5689–5698.
[20] R. Gautam and S. Latifi, "Comparison of simple missing data imputation techniques for numerical and categorical datasets," *Journal of Research in Engineering and Applied Sciences*, vol. 8, no. 1, pp. 468–475, 2023, doi: 10.46565/jreas.202381468-475.
[21] P. Buczak, J.-J. Chen, and M. Pauly, "Analyzing the effect of imputation on classification performance under mcar and mar missing mechanisms," *Entropy*, vol. 25, no. 3, 2023, doi: 10.3390/e25030521.
[22] M. B. Mohammed, H. S. Zulkafli, M. Adam, N. Ali, and I. Baba, "Comparison of five imputation methods in handling missing data in a continuous frequency table," *AIP Conference Proceedings*, vol. 2355, no. 1, p. 040006, 2021, doi: 10.1063/5.0053286.
[23] M. Dupuy, M. Chavent, and R. Dubois, "mdae : modified denoising autoencoder for missing data imputation," *ArXiv*, 2024, doi: 10.48550/arXiv.2411.12847.
[24] W. Hu, T. Wang, and F. Chu, "Fault feature recovery with wasserstein generative adversarial imputation network with gradient penalty for rotating machine health monitoring under signal loss condition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022, doi: 10.1109/tim.2022.3168898.
[25] U. I. Shabrina, R. Sarno, R. N. E. Anggraini, A. T. Haryono, and A. F. Septiyanto, "Sentiment analysis of presidential candidate debates from youtube videos," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, 2024, pp. 1–6, doi: 10.1109/AIMS61812.2024.10512640.
[26] K. E. ArunKumar, D. Kalaga, C. Mohan, S. Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends," *Alexandria Engineering Journal*, vol. 61, no. 10, pp. 7585–7603, 2022, doi: 10.1016/j.aej.2022.01.011.
[27] T. Shadbahr *et al.*, "The impact of imputation quality on machine learning classifiers for datasets with missing values," *Communications Medicine*, vol. 3, 2022, doi: 10.1038/s43856-023-00356-z.
[28] S. Robeson and C. Willmott, "Decomposition of the mean absolute error (mae) into systematic and unsystematic components," *PLOS ONE*, vol. 18, 2023, doi: 10.1371/journal.pone.0279774.
[29] J. Luo, G. Zhu, and H. Xiang, "Artificial intelligent based day-ahead stock market profit forecasting," *Computers and Electrical Engineering*, vol. 99, p. 107837, 2022, doi: 10.1016/j.compeleceng.2022.107837.
[30] A. T. Haryono, R. Sarno, R. N. E. Anggraini, and K. R. Sungkono, "Imputation missing stock prices using generative adversarial networks and attention mechanism," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, pp. 1–6, doi: 10.1109/ICTIIA61827.2024.10761233.

BIOGRAPHIES OF AUTHORS



Ulima Inas Shabrina received the master degree in Informatics from Institut Teknologi Sepuluh Nopember, Indonesia, in 2025. She currently join the research team in Institut Teknologi Sepuluh Nopember, focusing on the prediction of food supply and demand using machine learning models with panel data. Her research interests include machine learning, artificial intelligence, and generative models. She can be contacted at email: uishabrina@gmail.com.

Riyanarto Sarno earned his Ph.D. in 1992 and is currently a Professor in the Informatics Department at Institut Teknologi Sepuluh Nopember (ITS). He has authored more than five books and over 300 scientific papers, and in 2020, Stanford University recognized him among the top 2% of scientists worldwide. For the past five years, he has been conducting research in process mining. His areas of interest span machine learning, the internet of things, knowledge engineering, enterprise computing, and information management. He can be contacted at email: riyanarto@if.its.ac.id.

Ratih Nur Esti Anggraini received her doctoral degree in Engineering Mathematics from University of Bristol in 2021. She is a lecturer in Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. Her research interest includes text mining, artificial intelligence, and software engineering. She can be contacted at email: ratih_nea@if.its.ac.id.

Agus Tri Haryono was born in Balikpapan, Indonesia in 1988. He received a master's degree in informatics engineering from Institut Teknologi Sepuluh Nopember in 2023. He is pursuing a doctoral degree in computer science at the Institut Teknologi Sepuluh Nopember (ITS), Indonesia. His research interests include financial time series, text mining, and machine learning. He can be contacted at email: agustharyono24@gmail.com.