International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V74I1P124

Volume 74 Issue 1, 307-316, January 2026

Original Article

Performance Enhancement of GRU-Based Corn Yield Forecasting through MICE Imputation, PCA-Based Feature, and ELU Activation

Lyra K. Nuevas[1*], Marvee Cheska B. Natividad[2]

[1,2]*College of Computing Studies and Information Communication Technology, Isabela State University, Philippines.*

[*]*Corresponding Author : lyra.nueKvas_cyn@isu.edu.ph*

Received: 15 August 2025 Revised: 06 January 2026 Accepted: 08 January 2026 Published: 14 January 2026

Abstract *- Accurate corn yield forecasting is significant for farm agriculture to improve both farm production and productivity. In agriculture, forecasting is important to ensure food security. This study presents an improved Gated Recurrent Unit (GRU) forecasting approach that combines Multiple Imputation by Chained Equations (MICE), Principal Component Analysis (PCA), and Exponential Linear Unit (ELU) activation functions. This study is novel because it combines the PCA and ELU in a GRU architecture for agricultural forecasting. Previous study shows little to no work exploring this specific approach, or it is rarely seen in existing literature. MICE accurately imputes missing agronomic information, and PCA deals with multicollinearity and reduction of feature dimensions. This optimized input enhances gradient flow during training and mitigates the vanishing gradient issue common in deep recurrent models. Moreover, the application of the ELU activation stabilizes learning as it keeps small gradient values. The experiments showed that the model trained with dimensionality reduction with PCA, ELU activation, enhances performance with much higher accuracy than the baseline GRU models. The result produces fewer forecasting errors and consistent results with the actual yield values. These demonstrate that using the data imputation method combined with the ELU activation function enhances the performance of deep learning models in corn yield forecasting. This innovative solution gives farmers and agricultural planners managing a small farm or large operations the ability to make better decisions based on data.*

Keywords *- Corn yield forecasting, GRU, Multicollinearity, PCA, Vanishing gradient.*

1. Introduction

Corn serves as a worldwide essential food crop, which also functions as animal feed and provides raw materials for numerous industrial products that people use daily. Corn production directly influences food security, agricultural sustainability, and economic stability because developing countries depend on agriculture for their economic foundation. The most important and challenging part in agriculture is predicting crop yields because farmers are still using traditional methods in farming. The need to predict corn yields accurately becomes essential because it helps farmers make better decisions about farm operations and market placement of corn and supports the creation of agricultural governmental policies. Stakeholders will be able to evaluate their product availability by using yield forecasts together with other relevant data. Stakeholders must utilize reliable forecasts, along with additional elements, to achieve optimal resource management and mitigate risk from climate and environmental variability. Corn yield forecasting is complex due to multiple factors of soil characteristics, weather conditions, and crop management practices. Due to interactions among these and additional factors, corn yield forecasting, using historical data with missing values and multiple irregularities, is complicated. Commonly, linear regression methods and multiple regression analysis form the foundation of agricultural research related to crop yield forecasting. Across agricultural research, statistical techniques for these types of yield prediction, while having a generic method for implementation and interpretation, are restricted to using only linear or discrete data types based on the limitations of linear correlation models. Linear models for climate record statistical yield prediction will not provide accurate results for missing climate record values or missing yield history data. [1] The complexity and the extent of predicted yields associated with corn are compounded by the wide range of interrelated agronomic and environmental factors. Currently, the development of various machine learning and deep learning techniques has provided the agricultural industry with predictive models for the future yield of corn that provide a higher level of accuracy and dependability. Furthermore, given the ongoing rapid pace of transformation within the agricultural industry due to the current environment, the use

of deep learning models for yield prediction methods holds promise for providing optimal sustainable practices and sustainable farm operations. [2]

The agricultural sector shows growing interest in using multiple GRUs for different farming tasks, which include crop yield forecasting, weather prediction, natural language processing, and other applications of temporal data. [3] The GRU model achieves better performance results at lower computational cost when compared to LSTM and other models. LSTM models develop slowly due to the increase in the time it takes to train such models, and on the other hand, are classified as being an example of a large-scale model. [4] The performance of a GRU improves through advanced data pre-processing methods and activation function optimization techniques. MICE stand out as the most comprehensive method for addressing missing data with GRUs. [5]

To reduce the dimensions and lessen issues of multicollinearity, a common challenge in GRU, PCA was applied. PCA is noted to be an outstanding method to extract the most important patterns from complex datasets with minimal information loss and can thereby enhance the performance of the GRU models applied in high-dimensional cases. [6] The Exponential Linear Unit (ELU) activation function further helps in keeping the gradient flows stable and thus mitigates the vanishing gradient problem. Overall, GRU effectively facilitates learning complex patterns, hence giving better performance for models on agricultural data.

To fill this gap, the researchers introduce a new method of corn yield forecasting through a single pipeline, such as MICE-based imputation to handle missing data, PCA-based dimensionality reduction to remove complex features, and a GRU model with ELU activation functions to increase gradient flow. The difference between the method presented here and methods commonly used in previous studies is that many studies have focused only on the architecture of the yield forecasting model (LSTM or GRU) and/or the preprocessing methods used in their work, without considering the combined effect of all three methods as a whole of enhancing forecasting accuracy, the stability of reaching a predicted value, and the robustness of the model for predicting corn yields. Thus, by integrating the three methods into one framework, the researchers have created a novel framework for corn yield forecasting using real-world agronomic datasets containing both incomplete observations and features that are correlated.

2. Related Work

Over the years, machine learning methods have also been increasingly applied to agriculture because the models can interpret the relationship between complex patterns of inputs and outputs. [7] Forecasting crop yield is an important aspect for planning in the agriculture sector. A good forecasting method helps farmers decide early in preparation for the harvest period. [8] The data for traditional statistical methods is difficult to handle due to multiple characteristics, such as weather, soil, and farming practices. [1] Further research proves that deep learning models show a good performance in predicting crop yield. GRU for agricultural data since it can deal with the changing patterns over time [9]. Traditional statistical approaches, such as linear regression, have gained widespread acceptance in traditional crop yield forecasting methods due to their relative simplicity and interpretability. In reality, however, traditional modelling techniques do not sufficiently represent the complexities of agricultural production systems because they rely on assumptions of a linear relationship between farm production and climatic or environmental variables, assume that all relevant variables are independent of each other, and are based on a complete dataset. The assumptions of linearity, independence, and complete datasets do not hold for most agricultural production systems, and as such, linear regression models do not provide adequate representations of the complex nature of yield variation induced by climatic or environmental factors and associated agronomic interactions, nor do they yield very accurate or robust forecasting of crop yield potential.

2.1. Handling Missing Data in Agronomic Datasets

Agricultural researchers face common problems, such as missing data, incorrect measurements, equipment failure, and incomplete field sampling, which can complicate their work [10]. The most challenging aspect of farming is incomplete data, as most analytical tools do not function well when information is incomplete. It is essential to employ effective methods for addressing these issues. The model becomes more reliable and more accurate by replacing missing values. [11] Before the data is analyzed, the missing values will be addressed first. This can cause biased or inaccurate results if ignored. They may occur randomly, from systematic problems, or at other times, they happen while collecting or preprocessing data. [12] Missing data imputation has become important since real-world datasets almost always contain missing values. [13]

The common techniques for imputing missing values are KNN, MICE, and EM imputation; each technique uses different statistical methods. [11] The K-Nearest Neighbors algorithm performs well for classification and regression in agriculture by classifying new data by finding similarities to existing data. [14] As an instance-based learning method and a non-parametric approach, it deals with various agricultural datasets, such as crop yield forecasting and disease diagnosis. [15]. MICE is a reliable imputation technique that handles complex missing data patterns across different data types. [16] In contrast to single imputation methods that fill in missing values once, MICE creates multiple complete datasets, each with a different possible imputation for the missing values. [11] After reviewing the literature of the different authors, the researchers found that MICE consistently outperforms other imputation methods when it comes to preserving the original data distribution and variability.

2.2. GRU Models for Agriculture Time Series Forecasting

Farmers and researchers are increasingly using machine learning technologies in the agricultural sector to analyze agricultural data, create forecasts of crop yields, and design new agricultural methods that reduce the waste of resources. [17, 18] A popular machine learning model for analyzing weather information is the recurrent neural network model, including both Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) models. Many researchers have used Recurrent Neural Networks (RNNs), LSTMs, and GRUs to improve cancer prediction [19] and improve the way that analyzes the stock market using sentiment-aware sequential models [20], as well as improve the oil & gas production forecasting [21] and develop hydrological models for green roofs. [22] As more comparisons of GRU and LSTM have shown, there are both advantages and disadvantages associated with using GRU and LSTM models in time-series forecasting applications; GRUs typically train faster than LSTMs do in a lot of instances [23], whereas bidirectional GRU models have outperformed LSTMs in text classification. [24] While GRU offers advantages of faster training times and lower complexity due to fewer hyperparameters than LSTM, the benefits are seen more so when the models are trained on a small or limited number of datasets. [25] Therefore, as these advantages support the need for quick and efficient access to data to support productive agriculture, GRUs offer a viable model for farmers and researchers who have limited access to large volumes of data due to cost or other restrictions. GRUs have also been utilized effectively in previous studies to identify patterns within agricultural data to support accurate cropping practices and irrigation planning. [3] Additionally, GRU models are one class of deep learning technologies that have improved forecast accuracy for crop yield estimates, weather forecasts, and early detection of crop diseases. [26]

2.3. Forecast Modeling using PCA

Principal Component Analysis (PCA) is a way to take large amounts of data with correlated variables while preserving key patterns of complex datasets. [27] PCA in agricultural research helps by reducing dimensionality and multicollinearity, which improves model accuracy and makes it easier to understand and faster to run. By removing correlations and reducing dimensions, PCA makes the analysis simpler, and the results are easier to interpret. [28] This technique is useful for large farming datasets with many correlated features since it identifies the most important feature. [29]

2.4. Activation Functions Selection and Gradient Flow

The optimal choice of the activation function significantly influences the gradient flow in deep learning, such as the GRU-based model for time-series forecasting. [30, 31] For example, ReLU, Leaky ReLU, and ELU are good at ignoring the problem of vanishing gradient and letting the model train properly. [32, 33] Though ReLU is simple and boosts sparsity, neurons can always be permanently deactivated, and this is what was known as the "dying ReLU". [34] Recent studies have used smooth, non-saturating functions like Swish and ELU for better convergence and stability in training. Among them, ELU has the negative values that push the mean activations closer to zero for faster convergence and better performance compared to ReLU. [35]

2.5. Research Gap and Motivation

Existing research has established that utilizing recurrent neural networks, primarily Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, enhances the ability to predict crop yields accurately. However, most studies utilized just one preprocessing methodology, and there is generally insufficient exploration of how imputation quality, feature redundancy, and activation function parameters interact to affect (1) model training stability and (2) model generalization performance. Specifically, little exploration has been conducted regarding the benefit of jointly applying Multiple Imputation using Chained Equations (MICE) and Principal Component Analysis (PCA) as a single integrated preprocessing methodology within a GRU-based model for predicting corn yields, as well as optimizing activation functions. Therefore, to address these gaps, the goal of this research is to investigate the integrated contribution of MICE, PCA, and Exponential Linear Units (ELU) within a GRU framework.

3. Materials and Methods

The GRU architecture was selected for its lower complexity in terms of the number of model parameters compared to LSTMs, as well as its ability to converge more rapidly. This makes the GRU suitable for use with agricultural datasets that often have fewer data points. To handle missing data and maintain the relationship between predictor variables, MICE was used. PCA was used to reduce the number of predictor variables (dimensionality reduction) and help solve the problem of multicollinearity among variables, thereby increasing the stability of the model. The ELU activation function helps to improve the flow of gradients (gradient flow) and minimize problems related to vanishing gradients associated with deep recurrent networks. Model accuracy was measured using conventional evaluation metrics, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-Squared (R^2). The statistical significance was determined through paired t-tests to validate the performance of the baseline model.

3.1. Input Data and Preprocessing

The data collected from the Abuyog Experimental Station includes rainfall records collected over the past ten years and agronomic attributes of the corn yield. The dataset was examined before model development for missing values and inconsistencies, which are common in long-term agricultural data collection. Using MICE (Multiple Imputation by Chained Equations) methodology to fill in gaps from missing data to complete the datasets. The missing values were imputed

conditionally on the observed variables through the IterativeImputer's methodology to impute missing value variables iteratively onto the observed variable set. This methodology models each missing variable as a function of the other, remaining observed variables, in an iterative process. For this research study, the imputer was configured to a maximum of 10 iterations, with a fixed random seed assigned for the purposes of replicability. This methodology was chosen due to its capability of maintaining inter-relationships between the variables, and due to the reduced potential for bias that it offers when compared to single imputation models.

To assess the robustness of the imputation process, an artificial-missingness validation strategy was employed. A fixed proportion of observed values was temporarily masked, imputed using the same MICE configuration, and compared against the original values using standard error metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). This evaluation was performed exclusively on the masked entries to provide a valid assessment of imputation quality.

After imputing the missing data, 80% of the dataset was assigned to the training set and 20% to the testing set. Data augmentation techniques were then applied only to the training dataset to increase the diversity of unique examples available during model training while minimizing any potential for over-fitting to occur. The data in the testing set was not modified, which provides a true measure of how well the model will perform when presented with new, unseen data.

3.2. Feature Extraction using PCA

Principal Component Analysis (PCA) was used to reduce multicollinearity, improve efficiency, and decrease dimensionality within the dataset's features by applying PCA to them to analyze their relationships and interdependencies. The original raw dataset contains nine (9) input features, eight (8) agronomical and rainfall data, both measured on a monthly scale. The PCA identified seven (7) principal components that were able to explain 95% of the variance across all nine input variables; therefore, after applying PCA, the original feature space was reduced from nine (9) input variables to seven (7) principal components. These identified principal components became the basis for training the GRU (Gated Recurrent Unit) model, while retaining the target variable's original scale for interpretation. The selected variance threshold represents a trade-off between information preservation and dimensionality reduction.

A lower number of retained components may provide a loss of important agricultural information, and a higher number of retained components would make the reduced data less efficient for analysis and computation. This configuration has been selected to provide an optimal balance between the two by preserving most of the variability of the original dataset but increasing the stability of the training model and the efficiency of training the model. The target variable has been maintained in its original scale to continue allowing for ease of interpretation.

3.3. GRU Model Configuration

The GRU learns sequential data patterns effectively while addressing the vanishing gradient issue through its structure, similar to that of LSTM. A GRU provides a less complex architecture with fewer parameters than an LSTM, while the performance of the two models is quite similar in most time series forecasting applications. The GRU model for this research used an Exponential Linear Unit (ELU) activation function. The ELU activation function provides better performance than ReLU because it solves the vanishing gradient problem more effectively when used in deep neural networks. The ELU activation function produces negative values, which helps improve model learning dynamics and speed up convergence. The model training used the Adam optimization algorithm together with a Mean Squared Error (MSE) loss function.

3.4. Model Evaluation and Convergence Analysis

The model performance was evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 (Coefficient of determination). These metrics reflect different types of errors. The MAE was the average error size with respect to how far off the forecasts are. RMSE allows for heavier penalties on large deviations from actual values, while also looking at how sensitive the error is to large deviations, while R^2 indicates what proportion of the variance in corn yield is explained by the model itself. All three metrics that were used are common in the determination of yield forecasts; therefore, using all three provides a different perspective on the level of accuracy associated with model forecasting.

Throughout the entire training process, the training loss and the validation loss were compared against each other for convergence indication. The point of model convergence would be below the minimum value for validation loss; this is when model learning becomes stable and is at the least risk for over-fitting.

3.5. Statistical Significance Testing

To analyze the validity of the difference in performance observed between both GRU Models (with and without PCA), statistical significance testing for paired t-tests was performed at $\alpha = 0.05$. Pairing the absolute forecasting errors produced by the GRU Models evaluated on the same test dataset allowed for the evaluation of the statistically significant difference in forecasting error produced by each of the models used. The decision to compare the models' forecasting errors via a paired analysis was based on determining if the difference between the two model configurations is statistically significant and therefore not likely to have happened by chance.

4. Results and Discussion

4.1. Handling Missing Values with MICE

Multiple Imputation by Chained Equations (MICE) was used to address the agronomic data that is common to agriculture, which includes missing values, surpassing other methods of imputation. To determine what activation function for the GRU with MICE imputation, the four activation functions were tested: ELU, Leaky ReLU, ReLU, and Swish. To evaluate the forecasting accuracy, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 metrics were used. The convergence epoch was also used in this study to measure how quickly the model reaches stability during training and validation loss at convergence to measure how accurately the model performs once the training stabilizes. Table 1 indicates that the MICE-ELU performed best with a Mean Absolute Error (MAE) of 0.0007, Root Mean Squared Error (RMSE) of 0.0008, and an R^2 value of 0.9999. Also, it converged faster, achieving its best performance in only 4 epochs with a validation loss of 0.0011. While MICE-Leaky ReLU achieved a slightly lower validation loss of 0.0005, its error metrics were somewhat higher than MICE-ELU. MICE-Swish performed well in terms of R^2 but needed 6 epochs to converge-the longest among the three. Based on these results, MICE-ELU offers the best balance of accuracy, efficiency, and stability, making it the preferred choice for subsequent experiments with PCA-based dimensionality reduction and model comparisons.

Table 1. GRU Performance of activation functions using MICE

Activation	MAE	RMSE	R^2	Ep.	Val. Loss
ELU	0.0007	0.0008	0.9999	4	0.0011
Leaky ReLU	0.0010	0.0012	0.9999	5	0.0005
ReLU	0.0052	0.0070	0.9999	5	0.0003
Swish	0.0080	0.0107	0.9998	6	0.0001

4.1.1. Impact of MICE Imputation on Dataset Completeness

PCA reduced the original nine features to seven while keeping most of the important information. (a) shows the missing value patterns that were simulated before using Multiple Imputation by Chained Equations (MICE). The brown areas show that about 10% of the agronomic traits and rainfall data are missing. (b) shows the dataset after imputation with MICE. The uniform light shading means all missing values were filled in correctly. This shows how MICE can fill in missing values while keeping the data structure. Because the missing data pattern follows Missing at Random (MAR), MICE works well here. It utilizes the correlations between features to make unbiased imputations, making it suitable for the dataset. The heatmap shows that there is a non-uniform distribution of missingness in the agronomic and climatic variables and establishes the need for a strong imputation method. Because of this missingness pattern, it is more appropriate to use MICE, which will keep the relationships between the different variables instead of just simply removing them or imputing a mean.

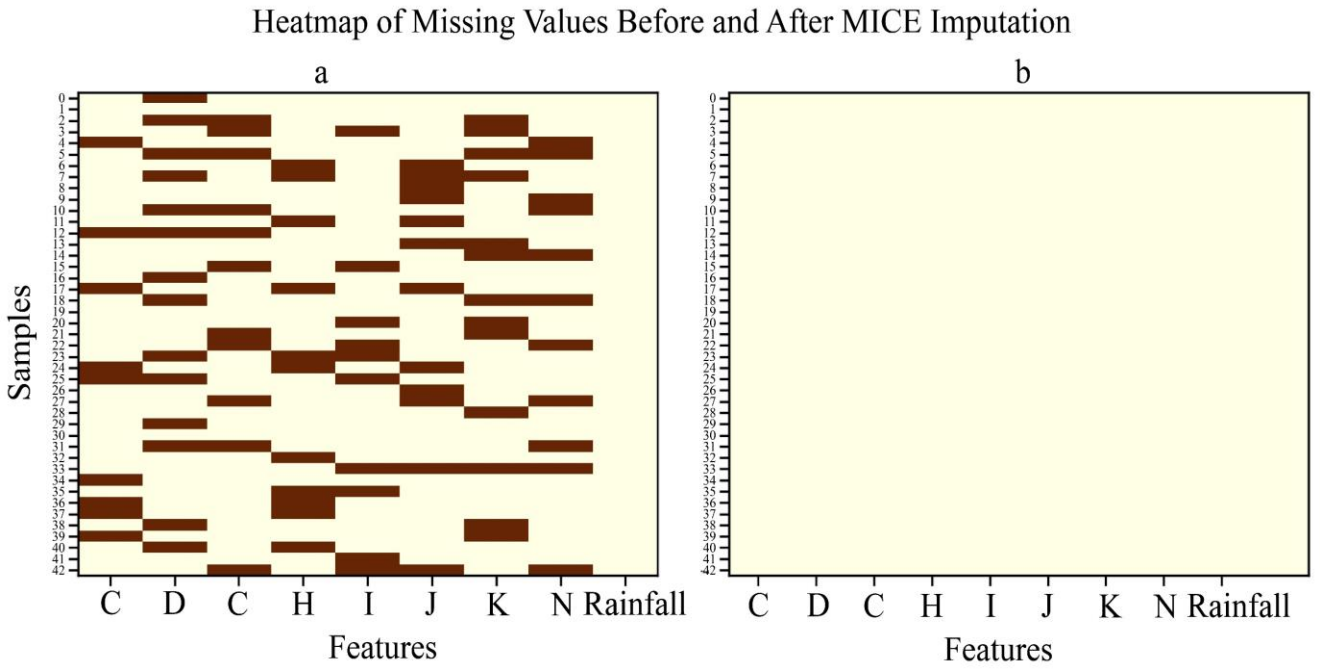

Fig. 1 Heatmap of missing values: (a) Simulated before MICE, and (b) After MICE imputation.

4.2. Dimensionality Reduction via PCA

After the correlation analysis, PCA was run on the normalized dataset (with imputed values by MICE) to address multicollinearity and reduce dimensions without losing important information. Seven components captured about 95% of the total variance, which contained sufficient information for further analysis. The loadings of the principal component analysis show that the original agronomic variables and original rainfall data, and their contributions to the first seven principal components.

The darker color represents higher absolute values, which indicates the feature has more influence on the principal component. Rainfall shows a strong contribution to PC4 because it has an absolute value of 0.90 on this component, as demonstrated in the Figure. The C, D, and G variables all have a strong loading in multiple principal components, indicating these variables both capture important patterns and differences in this dataset. This also supports using PCA to reduce the multicollinearity of the data while still allowing for meaningful agronomic information to be left in the dataset.

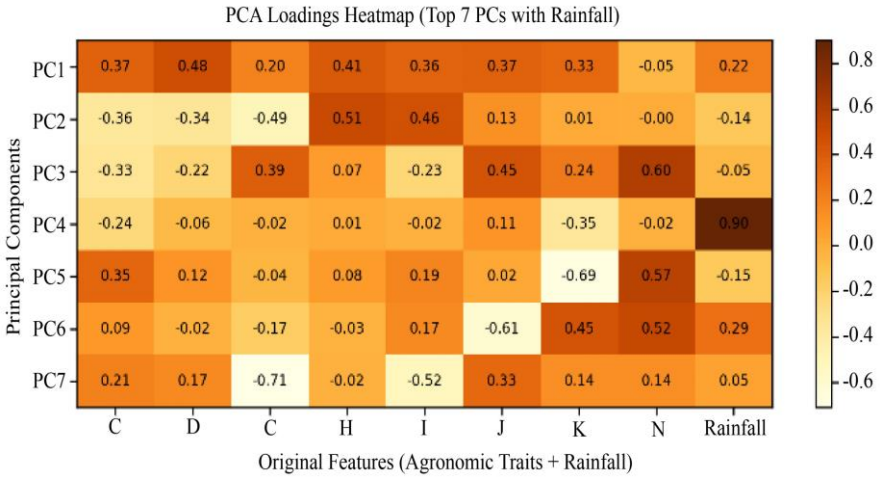

Fig. 2 PCA loading heatmap

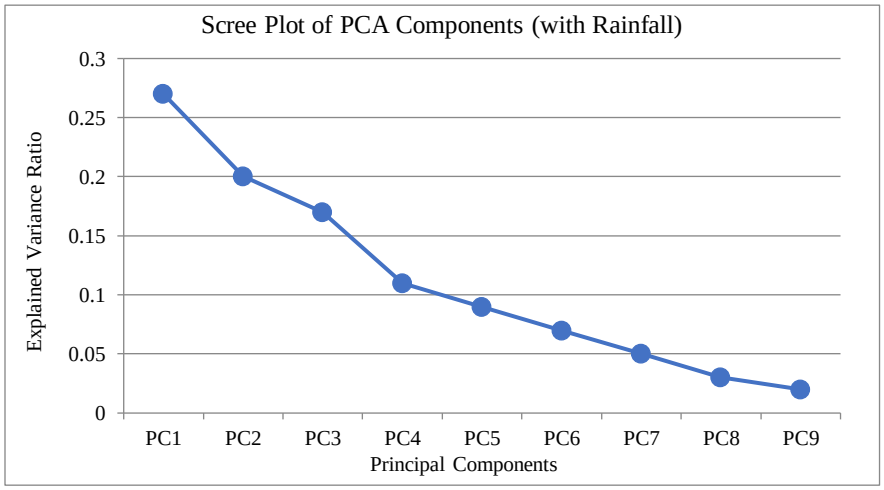

Fig. 3 Scree plot of PCA components with rainfall, showing explained variance ratio for each component

4.2.1. Scree Plot Analysis

A scree plot was created to find the best number of principal components for the corn yield dataset. The plot shows how much variance each component explains, helping identify where adding more components does not improve forecasting much. Number of Principal Components to retain 95% variance: 7. In Figure 3, a Scree Plot was created using PCA components for the MICE-imputed corn dataset. The first three principal components explain most of the variance: PC1 accounts for about 27%, PC2 for roughly 20%, and PC3 for around 17%. Later components contribute less to the overall data variability. Since the total variance reaches 95% by the seventh component, the first seven PCs were selected for further analysis. This reduced the dimensionality from nine input features to seven with minimal information loss, improving model efficiency while maintaining predictive accuracy. This was how seven principal components were chosen, as they represent an optimal trade-off between reducing the dimensionality of the data and keeping as much information about the agronomic variables intact.

4.2.2. Cumulative Variance Plot

Figure 4 shows the cumulative explained variance of the principal components from the MICE-imputed dataset with rainfall data. The cumulative variance rises quickly in the first few components, hitting around 64% with the first three components and going over 95% by the seventh component. This means the first seven components are enough to capture most of the original agronomic traits and rainfall data while reducing dimensionality. These results verify that PCA is capable of reducing dimensionality in the feature space very effectively without losing a considerable amount of information, thus making it suitable for creating optimized inputs for the GRU.

4.3. GRU Forecasting Accuracy before and after PCA

PCA reduced the nine original features to seven while keeping most of the dataset's variance. Using the original dataset, the GRU model had a Mean Absolute Error (MAE) of 142.06, a Root Mean Squared Error (RMSE) of 158.89, and an R^2 value of 0.1804.

After applying PCA, the results got much better: MAE dropped to 58.43, RMSE to 73.26, and R^2 jumped to 0.8258. MAE dropped down by about 58.8% and RMSE by 53.9%. A paired t-test confirmed the MAE reduction was significant ($t = 2.3369$, $p = 0.0476$), indicating PCA reduced redundancy and improved forecasting accuracy.

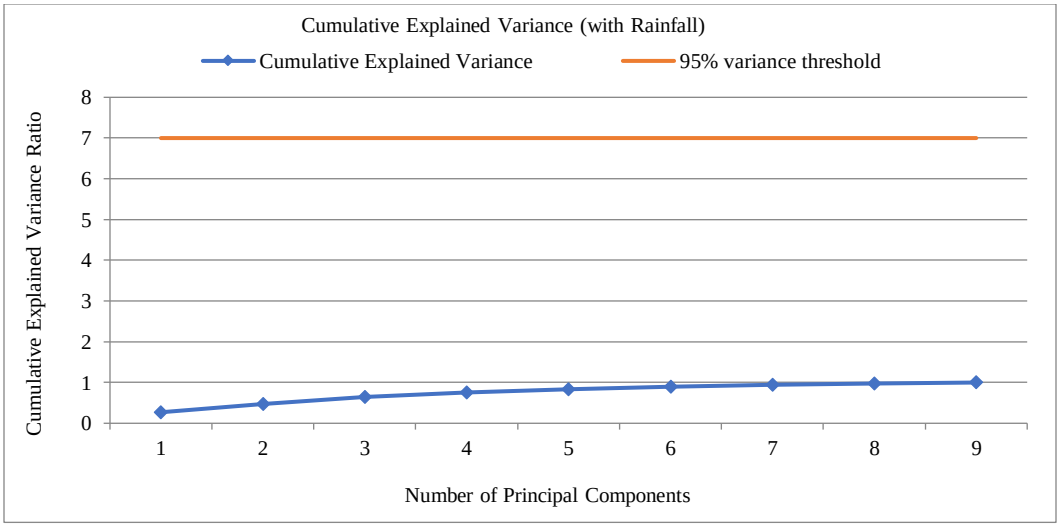

Fig. 4 Variance retention across principal components

Table 2. GRU model performance before and after PCA

Model	Features	MAE	RMSE	R^2 Score
GRU (Original Data)	9	142.06	158.89	0.1804
GRU (PCA Data)	7	58.43	73.26	0.8258

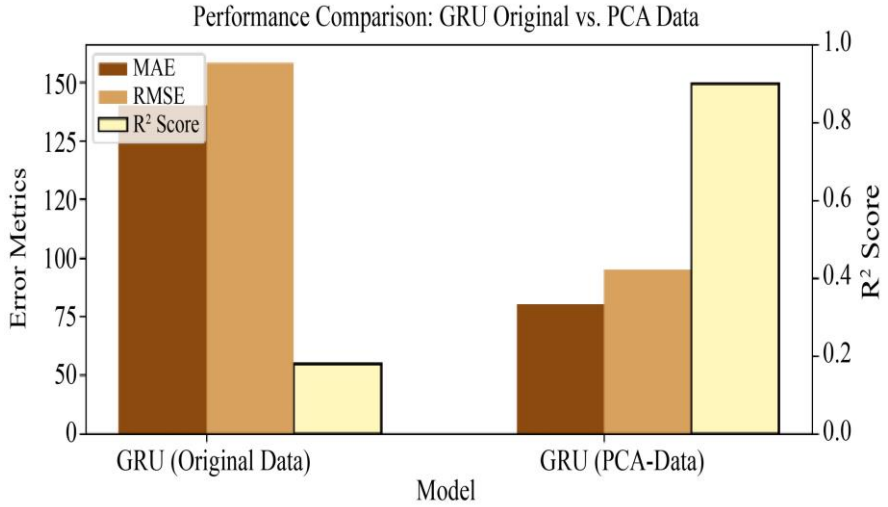

Fig. 5 GRU model performance with original versus PCA-reduced features

Figure 5 shows that PCA improved the model performance by reducing the features from nine down to seven, thus yielding MAE and RMSE, lower error rates, and a higher R^2 score, showing better forecasting accuracy.

The comparison of forecasting error and explanatory power for each model after introducing the use of PCA demonstrates that a large decrease in forecasting error and a large increase in explanatory power occurred as a result of the PCA process.

Thus, it can be concluded that through reducing dimensionality using PCA, the GRU's forecasting accuracy is enhanced due to the reduction of feature redundancy.

4.4. GRU Model Performance and Convergence Analysis PCA-Reduced Features

Figure 6 shows the training and validation loss curves for the GRU model using MICE, PCA, and ELU activation. Both curves drop sharply in the first few epochs and level off at low values after around 20 epochs. The training and validation losses follow each other closely, showing the model converges well and generalizes without overfitting. This shows that the GRU model is effective for corn yield forecasting. The convergence curves indicate that training loss and validation loss stabilized more smoothly and faster when using PCA features compared to all features. This indicates an enhancement in training stability through reduced overfitting caused by a more compact and well-conditioned input space.

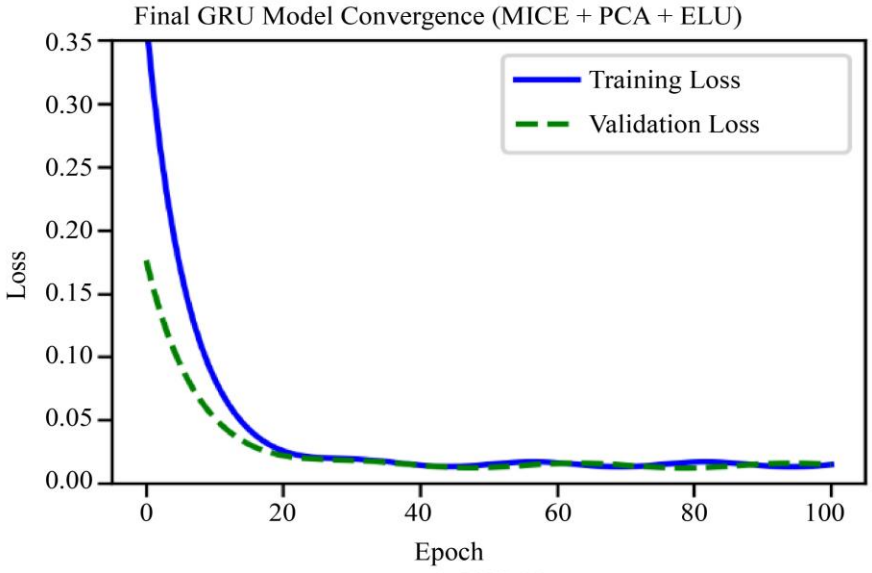

Fig. 6 GRU training behavior analysis

4.5. Discussion: Integrated Effect of MICE, PCA, and GRU

These findings suggest that different methods of preprocessing have a major impact on the accuracy and stability of predicting corn yields using GRU models. Data imputation using MICE will provide complete and accurate inputs into the model, and PCA will increase the predictive accuracy of the GRU model while also improving convergence stability through the reduction of multicollinearity. Thus, using a combination of data imputation and dimensionality reduction with GRU models provides a viable and effective method for forecasting agricultural time-series data, especially in situations where the dataset may be limited.

5. Conclusion

This research reveals that a novel forecasting approach that combines Gated Recurrent Units (GRU) with MICE, PCA, and ELU is successful in improving corn yield forecasting accuracy. GRU combined with PCA yielded a Mean Absolute Error (MAE) of 58.43, a Root Mean Squared Error (RMSE) of 73.26, and an R-squared of 0.8258. Compared to the original imputed features, the R-squared value of 0.1804 was much lower. The t-test for paired differences evaluating forecasting error obtained a t-statistic of 2.3369 and a p-value of 0.0476, indicating that this improvement was statistically significant at the 0.05 level.

The findings of the study support the claim that modeling corn yield forecasting using GRU with MICE, PCA, and ELU activation is a reliable and robust approach for accurate forecasting of corn yield, while improving the training process, as it reduces the vanishing gradient problem experienced in recurrent models.

Acknowledgments

The researchers sincerely thank Almighty God for giving us wisdom, strength, and good health through this rigorous academic challenge. This is indeed all possible through His guidance!

References

[1] Hongkun Fu et al., "Winter Wheat Yield Prediction using Satellite Remote Sensing Data and Deep Learning Models," *Agronomy*, vol. 15, no. 1, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sarmistha Saha et al., "Precision Agriculture for Improving Crop Yield Predictions: A Literature Review," *Frontiers in Agronomy*, vol. 7, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ming Dong, and Lukas Grumbach, "A Hybrid Distribution Feeder Long-Term Load Forecasting Method based on Sequence Prediction," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 470-478, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Thirupathi Kandadi, and G. Shankarlingam, "Drawbacks of LSTM Algorithm: A Case Study," *SSRN*, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Dimitris Bertsimas, Agni Orfanoudaki, and Colin Pawlowski, "Imputation of Clinical Covariates in Time Series," *Machine Learning*, vol. 110, no. 1, pp. 185-248, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ke Man et al., "Prediction of Rock Mass Classification in Tunnel Boring Machine Tunneling using the Principal Component Analysis (PCA)-Gated Recurrent Unit (GRU) Neural Network," *Deep Underground Science & Engineering*, vol. 3, no. 4, pp. 413-425, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Saeed Khaki, and Lizhi Wang, "Crop Yield Prediction using Deep Neural Networks," *Frontiers in Plant Science*, vol. 10, pp. 1-10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Djavan De Clercq, and Adam Mahdi, "Feasibility of Machine Learning-Based Rice Yield Prediction in India at the District Level using Climate Reanalysis Data," *arXiv Preprint*, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Khadija Meghraoui et al., "Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges," *Technologies*, vol. 12, no. 4, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Heikki Junninen et al., "Methods for Imputation of Missing Values in Air Quality Data Sets," *Atmospheric Environment*, vol. 38, no. 18, pp. 2895-2907, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[11] Christos Platias, and Georgios Petasis, "A Comparison of Machine Learning Methods for Data Imputation," *SETN 2020: 11th Hellenic Conference on Artificial Intelligence*, Athens, Greece, pp. 150-159, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Tlamelo Emmanuel et al., "A Survey on Missing Data in Machine Learning," *Journal of Big Data*, vol. 8, no. 1, pp. 1-37, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Khaled M. Fouad et al., "Advanced Methods for Missing Values Imputation based on Similarity Learning," *PeerJ Computer Science*, vol. 7, pp. 1-38, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Elife Ozturk Kiyak, Bita Ghasemkhani, and Derya Birant, "High-Level K-Nearest Neighbors (HLKNN): A Supervised Machine Learning Model for Classification Analysis," *Electronics*, vol. 12, no. 18, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Rajib Kumar Halder et al., "Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications," *Journal of Big Data*, vol. 11, no. 1, pp. 1-55, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Kirtan Jha et al., "A Comprehensive Review on Automation in Agriculture using Artificial Intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1-12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Juan Botero-Valencia et al., "Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives," *Agriculture*, vol. 15, no. 4, pp. 1-37, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan, "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61-69, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jovan Andjelkovic et al., "Sequential Machine Learning in Prediction of Common Cancers," *Informatics in Medicine Unlocked*, vol. 30, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Tej Bahadur Shahi et al., "Stock Price Forecasting with Deep Learning: A Comparative Study," *Mathematics*, vol. 8, no. 9, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Carlos Vega-Ortiz et al., "Decline Curve Analysis Using Machine Learning Algorithms: RNN, LSTM, and GRU," *Proceedings of the ARMA US Rock Mechanics/Geomechanics Symposium*, Atlanta, Georgia, USA, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Haowen Xie, Mark Randall, and Kwok-wing Chau, "Green Roof Hydrological Modelling with GRU and LSTM Networks," *Water Resources Management*, vol. 36, no. 3, pp. 1107-1122, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Shudong Yang, Xueying Yu, and Ying Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, Shanghai, China, pp. 98-101, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Hannan Asrawi, Ema Utami, and Ainul Yaqin, "LSTM and Bidirectional GRU Comparison for Text Classification," *Sinkron: Journal and Research of Informatics Engineering*, vol. 7, no. 4, pp. 2264-2274, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Yuchen Liu, Chaoxu Li, and Man Li, "Data-Driven Online State Prediction Method for the Traction Motors of Electric Multiple Units (EMUs)," *Sustainability*, vol. 17, no. 9, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] A.A. Mana et al., "Sustainable AI-Based Production Agriculture: Exploring AI Applications and Implications in Agricultural Practices," *Smart Agricultural Technology*, vol. 7, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Jiazhi Xia et al., "Revisiting Dimensionality Reduction Techniques for Visual Cluster Analysis: An Empirical Study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 529-539, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Aris Magklaras et al., "Enhancing Parameters Tuning of Overlay Models with Ridge Regression: Addressing Multicollinearity in High-Dimensional Data," *Mathematics*, vol. 12, no. 20, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Seungbum Koo, Dongik Shin, and Changhyuk Kim, "Application of Principal Component Analysis Approach to Predict Shear Strength of Reinforced Concrete Beams with Stirrups," *Materials*, vol. 14, no. 13, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Vladimír Kunc, and Jiří Kléma, "Three Decades of Activations: A Comprehensive Survey of 400 Activation Functions for Neural Networks," *arXiv Preprint*, pp. 1-107, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Alejandro Molina, Patrick Schramowski, and Kristian Kersting, "Padé Activation Units: End-to-End Learning of Flexible Activation Functions in Deep Networks," *Proceedings of the International Conference on Learning Representations (ICLR)*, pp. 1-17, 2020. [Google Scholar] [Publisher Link]

[32] Chigozie Nwankpa et al., "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," *arXiv Preprint*, pp. 1-20, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[33] Qi Zhang, and Teng Wang, "Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities," *Remote Sensing*, vol. 16, no. 8, pp. 1-47, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] Andrea Apicella et al., "A Survey on Modern Trainable Activation Functions," *Neural Networks*, vol. 138, pp. 14-32, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv Preprint*, pp. 1-14, 2016. [CrossRef] [Google Scholar] [Publisher Link]