

Food Loss Forecasting with Recurrent Neural Networks Using Gated Recurrent Units

Ru Poh Tan

*Tunku Abdul Rahman University
of Management and Technology,
Malaysia*
tanrp-wm21@student.tarc.edu.my

Siew Mooi Lim

*Tunku Abdul Rahman University
of Management and Technology,
Malaysia*
siewmooi@tarc.edu.my

Kuan Yew Leong

*A.I. System Research Co., Ltd.,
Japan*
kyleong@aisr.dev

Wei Zheng Goh

*Tunku Abdul Rahman University
of Management and Technology,
Malaysia*
gohwz-wm22@student.tarc.edu.my

Jun Kit Chaw

*Universiti Kebangsaan Malaysia,
Malaysia*
chawjk@ukm.edu.my

Siaw-Hong Liew

*Universiti Malaysia Sarawak,
Malaysia*
shliew@unimas.my

Shee Chia Lee

*Tunku Abdul Rahman University
of Management and Technology,
Malaysia*
leesheechia@tarc.edu.my

Abstract—Food loss forecasting is critical for global food security and sustainability. This study compares Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) architectures for predicting post-harvest food loss using Malaysian agricultural time-series data. Results demonstrate that GRU outperforms LSTM across all metrics, achieving 10.7% lower RMSE, 5.0% lower MAE, and 61.8% reduction in MAPE. GRU also exhibits superior computational efficiency with 23.9% fewer parameters and 19.5% faster inference. We release a reproducible setup that can be directly extended with Attention-GRU and seasonal-trend decomposition GRU (STL-GRU) to better capture long-range and seasonal dependencies. This work provides a practical foundation for next-generation models aimed at reducing losses and strengthening resilience in agri-food systems.

Keywords—RNN, GRU, food loss forecasting, deep learning

I. INTRODUCTION

Food loss represents one of the most critical long-term challenges to global food security and sustainability. The Food and Agriculture Organization of the United Nations (FAO) estimates that roughly one-third of food produced for human consumption is lost before reaching markets, driven by inadequate storage, poor logistics, and biophysical factors such as climate and pests [1]. These inefficiencies exacerbate hunger, cause substantial economic losses, and contribute to greenhouse gas emissions, making robust and accurate food loss forecasting a long-term priority for policymakers, researchers, and supply chain stakeholders [2].

Forecasting food loss is inherently complex, as losses arise from nonlinear interactions between biological perishability, climate variability, market volatility, and logistical constraints. Traditional time-series models such as ARIMA and exponential smoothing have been applied in agriculture [3], but they often fail to capture irregular seasonal patterns, long-range dependencies, and sudden shocks that characterize food loss dynamics. For instance, recent applications of ARIMA and ETS

to the wholesale market produce losses [4] demonstrate short-term utility but also highlight the limitations of linear approaches in providing durable predictive insights.

In contrast, deep learning models offer a more sustainable and powerful alternative for long-term forecasting. Recurrent Neural Networks (RNNs) and their gated variants, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have consistently demonstrated superior ability to capture nonlinear temporal dynamics in agricultural forecasting tasks such as yield prediction [5] and commodity price modeling [6]. The objective of this paper is to investigate the effectiveness of GRU-based recurrent neural networks for forecasting food loss. The key contributions are as follows:

- Develop and implement forecasting models using GRU and LSTM architectures for predicting post-harvest food loss from historical agricultural time-series data.
- Compare the predictive performance of GRU against standard LSTM baselines to determine improvements in accuracy and efficiency.
- Assess the capacity of GRU and LSTM models to generalize across different subsets of food loss data and capture temporal dynamics.

By rigorously evaluating the GRU architecture in the context of food loss prediction, this study contributes a reproducible and data-driven benchmark that not only informs future research and practical interventions aimed at mitigating losses within food systems, but also serves as a foundational reference for the next stage of our project. This work builds on our earlier exploration of the FAOSTAT dataset through clustering analysis of food security, waste, and loss in Malaysia [7], which highlighted the heterogeneity of loss drivers across agricultural systems. The present study advances that pipeline by establishing a baseline forecasting framework. In subsequent work, the GRU results reported here will be

systematically compared against enhanced architectures, including the Attention-GRU and Seasonal-Trend Decomposition GRU (STL-GRU), to assess their added value in capturing complex temporal dependencies and seasonal patterns in food loss data. Together, these interconnected efforts form the broader architecture of our research program, with the current baseline representing the starting point for progressively more sophisticated forecasting approaches.

II. LITERATURE REVIEW

A. Recurrent Neural Networks

The concept of RNNs was first introduced by Elman in 1990. RNNs capture sequential dependence by passing hidden states across time, which makes them a natural baseline for time-series forecasting tasks. They have been applied to agricultural contexts such as commodity price and yield prediction, where short-term temporal autocorrelation is significant. However, their ability to handle long-range dependencies is limited because of vanishing and exploding gradient problems [8]. Empirical evidence reinforces these shortcomings. For example, Manogna et al. showed that while RNNs can establish baseline forecasts for commodity prices, their accuracy degrades when dealing with longer and more complex temporal patterns, where gated models like LSTM and GRU achieve superior performance [6].

B. Advancements with Long Short-Term Memory

LSTM networks extend RNNs by introducing memory cells and gating mechanisms that allow information to be stored and selectively updated. This architecture directly addresses the vanishing gradient problem, enabling stable learning over extended sequences [9]. In agricultural forecasting, LSTMs have been widely applied to tasks involving seasonality and delayed effects, such as crop yield and commodity price prediction, where longer temporal memory improves accuracy. Studies comparing recurrent models consistently report that LSTMs outperform vanilla RNNs on noisy and nonlinear agricultural series. For example, [6] demonstrated that LSTM-based approaches achieved superior accuracy over RNNs in capturing long-horizon dependencies in commodity price data. Nonetheless, LSTMs are computationally more expensive and require large datasets for effective training.

C. Advancements with Gated Recurrent Units

GRU (4) were introduced as a streamlined alternative to LSTMs, combining the memory and gating concept into a simpler structure with fewer parameters [10]. The use of only update and reset gates reduces training time and computational cost, while still maintaining the ability to capture long-range dependencies in sequential data. This efficiency makes GRUs particularly attractive for agricultural applications, where datasets are often medium-sized, irregular, or heterogeneous. Comparative work has shown that GRUs achieve competitive forecasting accuracy with faster convergence compared to LSTMs. For instance, a study on the agricultural environment found that GRUs reached similar or better predictive performance than LSTMs while requiring fewer computational resources [11] and [12] shows that the performances of Attention-LSTM > GRU > LSTM. This finding inspired the potential benefits of developing an Attention-GRU model in

future research. Below are the equations for update gate (1), reset gate (2), candidate state (3) and final hidden state (4). At time step t , z_t and r_t represent the update and reset gate vectors, respectively. The W terms are the learnable weight matrices, while the b terms denote their corresponding biases. The input vector is x_t , and h indicates the hidden state vectors. The symbol $\hat{h}_t$ denotes the candidate hidden state, whereas h_t represents the updated or final hidden state. The operator $\circ$ indicates element-wise multiplication. For the initial step ($t = 0$), the hidden state is initialized as $h_0 = 0$. The update gate z and reset gate r determine how much information is preserved or discarded across time steps, enabling the model to effectively capture long-term dependencies.

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \quad (1)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \quad (2)$$

$$\hat{h}_t = \tanh(W_{\hat{h}}[r_t \circ h_{t-1}, x_t] + b_{\hat{h}}) \quad (3)$$

$$h_t = z_t \circ \hat{h}_t + (1 - z_t) \circ h_{t-1} \quad (4)$$

D. Research Gap: Food Loss Forecasting

Despite advances in agricultural forecasting, relatively few studies focus specifically on food loss prediction. The Food and Agriculture Organization (FAO) has developed global food loss estimation frameworks under Sustainable Development Goal (SDG) indicator 12.3.1a, emphasizing the urgency of reducing losses across supply chains [2]. Regional studies, such as clustering analyses in Malaysia, highlight the diverse drivers of loss across agricultural systems [7]. At the modeling level, Opara et al. [4] applied ARIMA and ETS to wholesale produce loss forecasting, demonstrating short-term utility but limited ability to capture nonlinear disruptions. The inherently noisy and heterogeneous nature of food loss data further challenges model robustness. These gaps underscore the need for stronger baselines using RNN, LSTM, and GRU models to improve predictive accuracy and provide actionable insights in this critical domain.

III. METHODOLOGY

A. Overall Workflow

This study employs a quantitative experimental design to evaluate the performance of recurrent neural network architectures, LSTM and GRU, for forecasting post-harvest food loss. Fig. 1 below shows the structured workflow. Each model is trained and validated on the same dataset to ensure a fair comparison of predictive accuracy and computational efficiency. Note that this forecasting represents aggregated results across all items, rather than individual item-level forecasts.

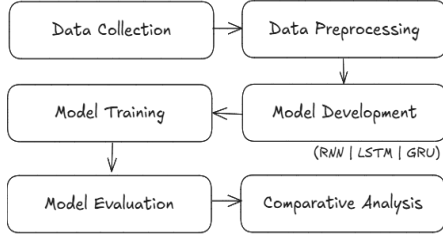


Fig. 1 Overall Workflow

1. Data Collection

The dataset comprises historical agricultural time-series records of food loss across selected commodities, sourced from the FAO and agricultural reports. It captures post-harvest losses at production and distribution stages for 10 years in Malaysia only agriculture. Challenges include irregular time intervals, missing values, and heterogeneous scales across commodities, necessitating thorough preprocessing before model application.

2. Preprocessing and Exploratory Data Analysis

Preprocessing involved filtering for primary commodities, imputing missing values with zeros for non-reported figures, and conducting sanity checks to validate production balances between Quality Control Loss (QCL) and Supply and Utilization Account (SUA). Outliers were detected using IQR and z-scores but retained, as they reflected genuine agricultural shocks. Feature engineering introduced per-capita availability, loss ratios, domestic availability, Self-Sufficiency Ratio (SSR), Import Dependency Ratio (IDR), and aggregated commodity indicators. Numerical features were transformed to reduce skewness and scaled using log transformation, Min-Max scaling, Robust scaling, or standardisation, depending on their statistical properties.

Exploratory Data Analysis (EDA) was performed to validate data consistency and reveal structural patterns in Malaysia's food system. Distribution analysis before and after scaling (Fig. 2) demonstrates the effect of normalisation in reducing skewness.

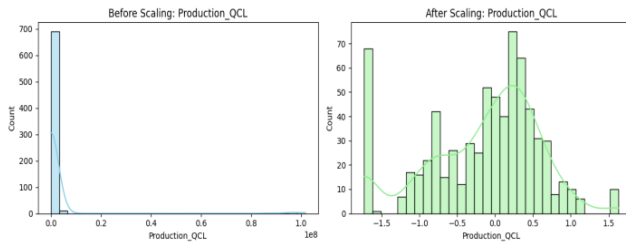


Fig. 2 Normalised distribution

Commodity-level rankings of the top ten crops by harvested area, production, and loss were generated to illustrate the dataset's composition. As shown in Fig. 3, the commodity rankings are included solely to illustrate the composition of the dataset and do not influence the forecasting task, which is based on aggregated values across all commodities.

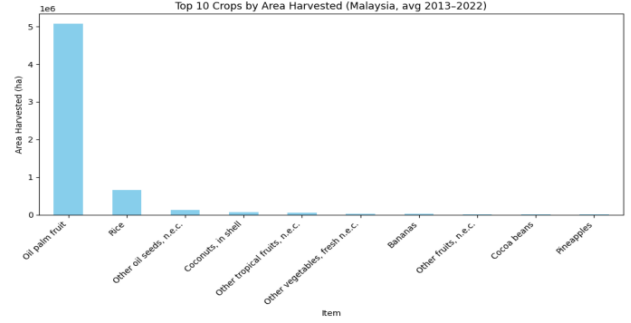


Fig. 3 Top 10 Crops by Area Harvested

Temporal line plots revealed clear long-term production and loss trends between 2013–2022, showing fluctuations in key commodities that reflect structural shifts in Malaysia's agricultural system. Pair plots further confirmed the expected positive relationship between yield and production, while also exposing weaker or inconsistent correlations for secondary variables. A parallel coordinate plot of the top commodities enabled multi-dimensional comparison across area harvested, production revealed the production efficiency of certain crops, and the correlation heatmap (Fig. 4) revealed interdependencies among features useful for selection. Collectively, these analyses ensured a reliable and feature-rich dataset for model development.

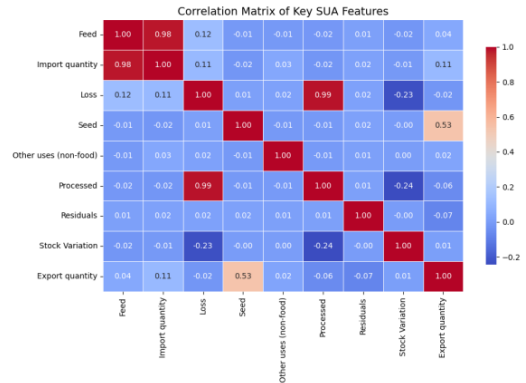


Fig. 4 Correlation heatmap

3. Model Development

Two recurrent neural architectures were implemented for comparative analysis:

- LSTM: Extends RNN with memory cells and gating mechanisms to mitigate vanishing gradient issues and capture long-term dependencies.
- GRU: Simplifies LSTM gating into update and reset gates, achieving comparable accuracy with reduced parameter complexity and faster training.

Both models adopt an identical architecture and hyperparameter configuration, with the only distinction being the replacement of GRU cells with LSTM cells. They were implemented using the TensorFlow/Keras Sequential API and consist of two recurrent layers followed by dense and dropout layers. Specifically, the LSTM model includes two stacked

LSTM layers (96 and 32 units, respectively), each regularized with L2 ($\lambda = 0.001$) and interleaved with dropout layers (rate = 0.3) to mitigate overfitting. The dense and output layers are configured identically to the GRU counterpart, featuring a 32-unit ReLU layer and a single linear output neuron for continuous loss prediction. This architectural parity ensures that any performance differences observed between the GRU and LSTM models arise solely from their respective gating mechanisms, rather than from structural or training disparities.

4. Model Training and Evaluation Metrics

The training of both GRU and LSTM models was conducted under identical configurations to ensure a fair and consistent performance comparison. Each model was trained using approximately 70% of the total sequences as training data, 15% for validation, and 15% for testing, maintaining a strict temporal split to avoid data leakage. The models employed a batch size of 32 and were trained for a maximum of 200 epochs, with an initial learning rate of 0.001, L2 regularization ($\lambda = 0.001$), and a dropout rate of 0.3 to mitigate overfitting. Gradient clipping was applied with a clip norm of 1.0 to maintain training stability. The Adam optimizer was employed with a learning rate of 0.001, optimizing the Mean Squared Error (MSE) loss function, while MSE and Mean Absolute Error (MAE) were tracked as performance metrics. Training incorporated multiple callbacks for robust learning control: Early Stopping monitored the validation MAE with a patience of 20 epochs and automatically restored the best-performing weights; Learning Rate Reduction on Plateau halved the learning rate (factor = 0.5) when validation MAE plateaued for 10 epochs, with a minimum learning rate threshold of $1e-7$; and a Model Checkpoint callback saved only the best model based on the lowest validation MAE to 'gru_model_best_scaled.keras'. Validation was performed using an explicitly separated validation set (X_{val}, y_{val})—not by random split—to preserve temporal integrity, with shuffling applied only within the training set to enhance generalization while avoiding data leakage.

The evaluation focuses on both scaled and original performance metrics, including:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Coefficient of Determination (R^2)
- Mean Absolute Percentage Error (MAPE)

Additionally, inference efficiency and relative improvements were also assessed.

IV. RESULTS AND DISCUSSION

A. Model Evaluation Results

On the scaled metrics (Table I), both models showed comparable performance, with GRU achieving slightly lower MAE (0.220 vs 0.232). While Fig. 5 presents the evaluation metrics on the original scale. The GRU model achieved

superior predictive accuracy across all metrics: RMSE of 0.219 (10.7% improvement over LSTM's 0.245), MAE of 0.112 (5.0% improvement over LSTM's 0.118), and R^2 of 0.805 compared to LSTM's 0.736. Most notably, GRU demonstrated substantially lower MAPE (5.97%) compared to LSTM (15.64%), indicating 61.8% reduction in percentage error.

TABLE I. EVALUATION PERFORMANCE (SCALED)

Metrics	LSTM	GRU
MSE	0.2310	0.1844
MAE	0.2319	0.2203

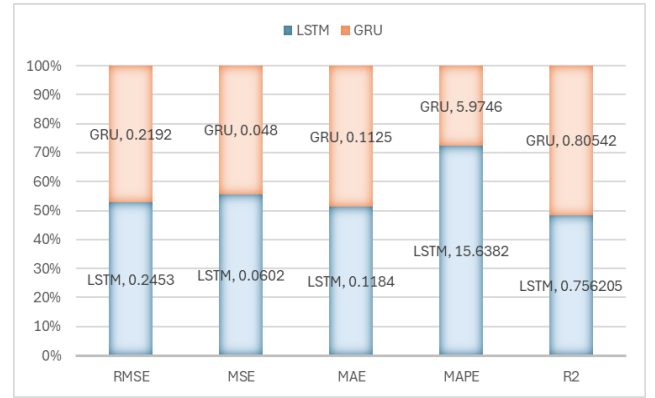


Fig. 5 Evaluation Performance for Original Scaled.

Fig. 6 below reveals that GRU exhibits better generalization, with validation MAE of 0.087 compared to LSTM's 0.107, suggesting reduced overfitting.

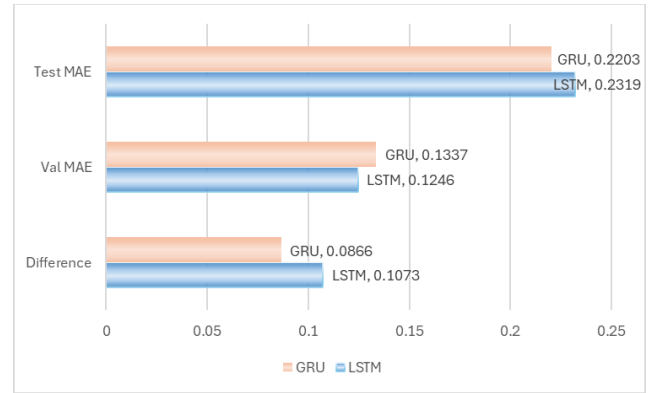


Fig. 6 Comparison with Validation

Fig. 7 below illustrates the forecast comparisons across multiple time windows, highlighting the differences between the Actual, GRU, and LSTM predictions.

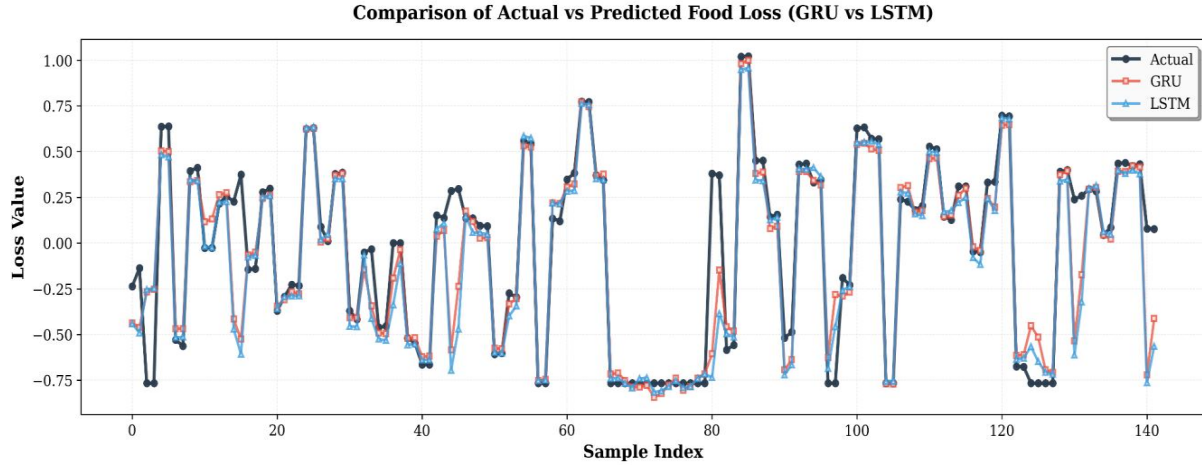


Fig. 7 Actual vs GRU vs LSTM: Multiple Time Windows.

B. Computational Efficiency

Table II demonstrates GRU's computational advantages. Despite slightly longer training time (0.54 min vs 0.47 min), GRU achieved 23.9% parameter reduction (46,401 vs 60,993), 23.0% smaller model size (0.57 MB vs 0.74 MB), and 19.5% faster inference per sample (2.46 ms vs 3.05 ms). This efficiency advantage is critical for deployment in resource-constrained agricultural monitoring systems.

TABLE II. COMPUTATIONAL EFFICIENCY COMPARISON.

Metrics	LSTM	GRU
Training Time	0.47 min	0.54 min
Total trainable parameters	60,993	46,401
Model size on disk	0.74 MB	0.57 MB
Total inference time	0.4333 sec	0.3489 sec
Average inference time per sample	3.0513 ms	2.4571 ms

C. Convergence and Learning Stability

Both networks exhibit smooth convergence; however, the GRU model demonstrated faster loss stabilization and lower validation loss variance, implying more stable learning dynamics. GRU demonstrated a convergence rate of 47.94% compared to LSTM's 58.17%. While LSTM showed faster convergence, GRU's final performance metrics were superior, suggesting that the slower convergence led to better optimization and reduced overfitting. Fig. 8 demonstrates the training and validation loss curves for (a) GRU and (b) LSTM.

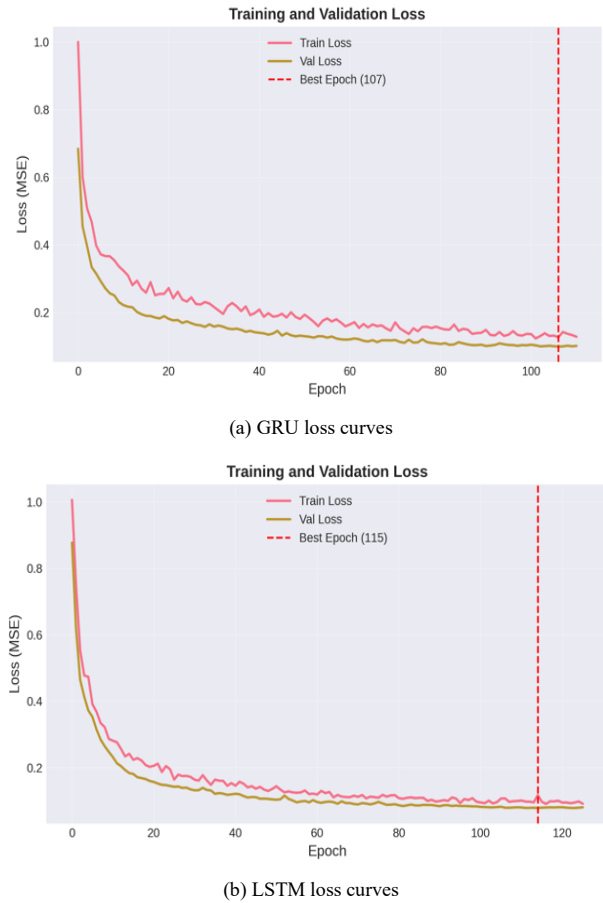


Fig. 8 Training and Validation Loss Curves for (a) GRU and (b) LSTM.

V. CONCLUSION AND FUTURE WORK

This study evaluated and compared GRU and LSTM architectures for time-series forecasting of food loss prediction. Experimental results demonstrate that GRU consistently outperforms LSTM in accuracy and generalization, achieving lower RMSE and MAE on both scaled and original test sets. Specifically, GRU recorded an RMSE of 0.219 (10.7% improvement), MAE of 0.112 (5.0% improvement), and R^2 of 0.805 compared to LSTM's 0.736, while reducing MAPE by 61.8% (5.97% vs. 15.64%). The integration of L2 regularization and dropout further mitigated overfitting, resulting in smoother convergence and stable learning dynamics.

Beyond accuracy, GRU also demonstrated superior computational efficiency, training with 23.9% fewer parameters, a 23.0% smaller model size, and 19.5% faster inference per sample. Despite slightly longer training time, GRU achieved better optimization stability, confirming its suitability for real-time or resource-constrained agricultural forecasting applications. Collectively, these results confirm that the GRU architecture provides a more compact and efficient solution for sequence modeling without compromising predictive accuracy.

By rigorously evaluating the GRU architecture in the context of food loss prediction, this study establishes a reproducible, data-driven benchmark that informs both academic research and practical interventions aimed at minimizing agricultural losses. It also builds upon our earlier exploration of the FAOSTAT dataset through clustering analysis of food security, waste, and loss in Malaysia [7], which revealed heterogeneous loss drivers across agricultural systems. The present work extends that analytical pipeline by defining a robust baseline forecasting framework.

In future work, this GRU baseline will be systematically enhanced through memory-augmented architectures, including Attention-GRU and Seasonal-Trend Decomposition GRU (STL-GRU) models, to further strengthen the ability to capture long-term dependencies, seasonal dynamics, and contextual variations in food loss data. These advancements represent the next phase of our broader research program, where the current GRU benchmark serves as the foundational stage toward

progressively more intelligent and adaptive forecasting models for sustainable food system management.

ACKNOWLEDGMENT

This work is funded by Tunku Abdul Rahman University of Management and Technology (TAR UMT) through the Internal Research Grant: UC/I/G2025-00164.

REFERENCES

- [1] H. Breewood, W. Fraanje, and T. Garnett, "What is food loss and food waste?," *Food Climate Research Network*, Mar. 2019. doi: 10.56661/f98ed9f6.
- [2] Charlotte Taglioni, José Rosero Moncayo, and Carola Fabi, *Food loss estimation: SDG 12.3.1a data and modelling approach*. in 23-39. FAO, 2023. doi: 10.4060/cc9173en.
- [3] Rob J. Hyndman and George Athanasopoulos, *Forecasting: principles and practice*, 3rd edition. OTexts, Melbourne, Australia, 2021. Accessed: Sep. 25, 2025. [Online]. Available: <https://otexts.com/fpp3/>
- [4] I. K. Opara, Y. Silue, U. L. Opara, and O. A. Fawole, "Forecasting of postharvest fresh produce waste at the wholesale level using time series models," *Agric. Food Secur.*, vol. 14, no. 1, p. 24, Sep. 2025, doi: 10.1186/s40066-025-00550-3.
- [5] J. Wang, P. Wang, H. Tian, K. Tansey, J. Liu, and W. Quan, "A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables," *Comput. Electron. Agric.*, vol. 206, p. 107705, Mar. 2023, doi: 10.1016/j.compag.2023.107705.
- [6] R. L. Manogna, V. Dharmaji, and S. Sarang, "Enhancing agricultural commodity price forecasting with deep learning," *Sci. Rep.*, vol. 15, no. 1, p. 20903, Jul. 2025, doi: 10.1038/s41598-025-05103-z.
- [7] E. Chen Sheng Hii, S. M. Lim, S. X. Loo, S. K. Yeap, and C. Y. Tan, "Clustering Analysis of Food Security, Waste and Loss: Malaysia Agricultural Insights," *JOIV Int. J. Inform. Vis.*, vol. 8, no. 3–2, p. 1892, Nov. 2024, doi: 10.62527/joiv.8.3-2.3152.
- [8] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," Oct. 17, 2015, *arXiv:1506.00019*. doi: 10.48550/arXiv.1506.00019.
- [9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [10] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," Sep. 03, 2014, *arXiv:1406.1078*. doi: 10.48550/arXiv.1406.1078.
- [11] S. Chen, B. Li, J. Cao, and B. Mao, "Research on Agricultural Environment Prediction Based on Deep Learning," *Procedia Comput. Sci.*, vol. 139, pp. 33–40, Jan. 2018, doi: 10.1016/j.procs.2018.10.214.
- [12] Z. Xiong, X. Guo, D. Zhang, S.-P. Tseng, Y. Ji, and X. Sun, "Crop Yield Prediction based on Attention-LSTM Model," in *2024 International Conference on Orange Technology (ICOT)*, Dec. 2024, pp. 1–4. doi: 10.1109/ICOT64290.2024.10936959.