Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol 6, No.2, September 2023, pp. 208 – 219
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Harnessing The Power of Stacked GRU For Accurate Weather Predictions

[1]Mohammad Diqi, [2]Ahmad Wakhid, [3]I Wayan Ordiyasa, [4]Nurhadi Wijaya, [5]Marselina Endah Hiswati
[1,3,4,5]Department of Informatics, Universitas Respati Yogyakarta, Yogyakarta, Indonesia
[2]School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China
Email: [1]diqi@respati.ac.id, [2]wakhidahmad26@gmail.com, [3]wayan@respati.ac.id, [4]nurhadi@respati.ac.id, [5]marsel.endah@respati.ac.id

Article Info
Article history:
Received Jun 19th, 2023
Revised Jul 24th, 2023
Accepted Aug 30th, 2023

Keyword:
Model Performance
Prediction Accuracy
Stacked Gated Recurrent Unit
Time-Series Data
Weather Forecasting

ABSTRACT
This research proposed a novel approach using Stacked GRU (Gated Recurrent Unit) models to address the problem of weather prediction and aimed to improve forecasting accuracy in sectors like agriculture, transportation, and disaster management. The key idea involved leveraging the temporal dependencies and memory management capabilities of Stacked GRU to model complex weather patterns effectively. Comprehensive data preprocessing ensured data quality and fine-tuning of the model architecture and hyperparameters optimized performance. The research demonstrated the Stacked GRU model's effectiveness in accurately forecasting temperature, pressure, humidity, and wind speed, validated by low RMSE and MAE scores and high R2 coefficients. However, challenges in forecasting humidity and a percentage discrepancy in wind speed predictions were observed. Overfitting and computational complexity were identified as potential limitations. Despite these constraints, the study concluded that the Stacked GRU model showed promise in weather forecasting and warranted further refinement for broader applications in time-series prediction tasks.

Corresponding Author:
Mohammad Diqi,
Department of Informatics,
Universitas Respati Yogyakarta,
Jl. Laksda Adisucipto KM 6,3 Caturtunggal Depok Sleman, Yogyakarta, Indonesia.
Email: diqi@respati.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v6i2.24769

1. INTRODUCTION

Weather prediction is important due to its broad impact on agriculture, transportation, and disaster management sectors [1]. Accurate forecasts contribute to economic improvements by facilitating planning and decision-making [2]. Various technological advancements are required for this intricate task, including innovations in computation and measurement systems [3]. Meteorologists have long faced the challenge of accurate weather predictions [4]. Recent years have seen increased efforts to develop unified models and frameworks for weather and climate predictions, emphasizing the need for consistency across time scales [5]. Forecasting specific weather phenomena like rainfall and windstorms is necessary for applications such as agriculture and building assessments [2], [6]. Researchers have explored advanced techniques like data mining, deep learning, and fuzzy logic to enhance weather prediction model accuracy and reliability [7], [8]. Integrating weather predictions into decision-making can optimize resource allocation and preventive actions [6].

Accurate weather prediction significantly impacts various sectors, such as agriculture, energy, and transportation. In agriculture, precise weather forecasts help farmers make informed decisions regarding planting, irrigation, and pest control, leading to improved crop yields and resource management [9]. For the energy sector, accurate weather prediction is crucial for optimizing the generation and distribution of renewable energy sources like solar and wind power, enabling efficient operations and planning [10]. In transportation, weather forecasts aid in route planning, scheduling, and safety measures, reducing the risk of accidents and improving efficiency [11]. Additionally, accurate weather prediction is vital in disaster management, allowing for timely evacuation and preparation for extreme weather events [12]. Overall, accurate weather prediction enhances decision-making processes, minimizes risks, and optimizes resource allocation in these sectors, leading to economic and societal benefits [13].

Current methodologies used in weather prediction include ensemble prediction methods, machine learning models, deep learning models, and numerical weather prediction (NWP) models. Ensemble prediction methods generate multiple predictions using different parameter values or initial conditions to account for uncertainty [14]. Machine learning models, such as artificial neural networks (ANNs), have been used to improve the accuracy of weather forecasts [15], [16]. Deep learning models, such as long short-term memory (LSTM) networks, have shown promise in capturing temporal dependencies in weather data [15]. NWP models use mathematical equations to simulate atmospheric processes and predict weather conditions [17]. However, these methodologies have limitations. Terministic models have inherent uncertainty due to the complexity of weather systems. Ensemble methods can be computationally expensive and require large amounts of data [14]. Machine learning models may lack interpretability and struggle with extreme weather events [18]. Deep learning models require large amounts of training data and can be computationally intensive [19]. NWP models rely on accurate initial conditions and parameterizations, which can introduce errors [20]. Additionally, the temporal averaging of covariate data in niche modeling can limit predictive capacity for species affected by short-term environmental changes [20]. While these methodologies have advanced weather prediction, there are still challenges regarding accuracy, computational efficiency, interpretability, and handling extreme events.

For several reasons, exploring new machine learning methodologies for weather prediction is vital. Firstly, traditional NWP models have limitations, and new approaches can help overcome these limitations [20]. Secondly, machine learning techniques have shown promise in improving weather prediction accuracy and capturing complex patterns in weather data [21], [22]. Thirdly, exploring new methodologies can lead to advancements in extreme weather event forecasting, which is crucial for disaster management and preparedness [23]. Machine learning models can also provide insights into underlying weather prediction and climate diagnosis mechanisms [21]. Finally, developing new methodologies can enhance the accuracy and effectiveness of weather forecasts, benefiting various sectors such as agriculture, transportation, and energy [24]. By exploring new machine learning methodologies, we can improve the accuracy, efficiency, and understanding of weather prediction, leading to better decision-making and societal benefits.

A Gated Recurrent Unit (GRU) is a type of recurrent neural network (RNN) that was introduced as an alternative to the more complex LSTM units [25]. GRUs are designed to handle long-term dependencies in sequential data [26]. They operate by utilizing gating mechanisms, including an update gate and a reset gate, to control the flow of information within the network [27]. The update gate determines how much of the previous hidden state should be retained, while the reset gate controls how much of the previous hidden state should be forgotten [28]. This gating mechanism allows GRUs to selectively update and forget information and capture relevant patterns in the input sequence [27].

Compared to other RNN models, GRUs have a simpler architecture with fewer gating mechanisms, making them computationally efficient. GRUs have been found to perform comparably to LSTMs in various machine learning tasks, including sequence prediction tasks [27]. They have shown promise in applications such as speech recognition [26], splice site prediction [29], time series forecasting [30], and dynamic risk prediction in healthcare [31]. Additionally, GRUs have been explored with other models, such as convolutional neural networks (CNNs), to improve performance in tasks like electric load forecasting [32].

The GRU model was chosen over other models for this analysis due to its advantages in weather prediction. GRU models have been widely used in various domains, including electric load forecasting [33], solar irradiance forecasting [34], precision agriculture [35], carbon dioxide concentration prediction [36], traffic prediction [37], landslide displacement prediction [38], wind speed and temperature forecasting [39], wildfire detection [40], and solar radiation prediction [41]. The advantages of GRU models in weather prediction include their simplicity and ease of implementation [33], ability to capture long-term dependencies in sequential data [34], improved performance with attention mechanisms [33], and computational efficiency compared to other recurrent neural network models [37]. GRU models have shown promising results in accuracy, prediction performance, and efficiency in various weather prediction tasks, making them a suitable choice for this analysis.

A GRU handles long-term dependencies in sequential data for weather forecasting through its gating mechanisms. The update and reset gates in a GRU control the flow of information, allowing the model to selectively update and forget information from the previous hidden state [34]. This enables the GRU to capture relevant patterns and dependencies in the input sequence, facilitating the modeling of long-term dependencies [26].

Stacking multiple GRU layers improves model performance by allowing for the extraction of increasingly complex features from the input data. Each additional layer enables the model to learn higher-level sequential data representations, improving prediction accuracy [42]. The stacked GRU architecture captures hierarchical patterns and dependencies in the data, enhancing the model's ability to capture long-term dependencies and make accurate predictions [43].

The advantages of GRUs in weather prediction include their ability to handle long-term dependencies [43], computational efficiency [26], and competitive performance in sequence prediction tasks [26]. GRUs have been successfully applied in various weather forecastings tasks, such as solar irradiance [34], wind speed and temperature forecasting, and long-term weather forecasting [44]. Their ability to capture complex patterns and handle long-term dependencies makes GRUs suitable for weather prediction tasks.

The expected benefits of using a Stacked GRU model for weather prediction using the Denpasar Weather Data include improved prediction accuracy, capturing long-term dependencies in weather patterns, and efficient computation. Stacked GRU models have been shown to outperform other models in various prediction tasks, such as uncertainty estimation [45], sentiment analysis [46], parking occupancy prediction [47], traffic prediction [37], precipitation forecasting [12], and disease prediction [48]. The advantages of using Stacked GRU models include their ability to handle long-term dependencies [26], capture complex patterns in sequential data [22], and improve prediction accuracy [42]. By leveraging the strengths of GRU models, the Stacked GRU model is expected to provide accurate and efficient weather predictions using the Denpasar Weather Data.

2. RESEARCH METHOD

2.1. Gated Recurrent Unit (GRU)

The GRU is a recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data. It achieves this through the use of gating mechanisms that control the flow of information within the network.

Mathematically, the GRU layer consists of several equations to update and control its internal states. Let us denote the input to the GRU layer at timestep t as x_t, the hidden state at the previous timestep as h_{t-1}, and the updated hidden state at timestep t as h_t.

The update gate z_t determines how much of the previous hidden state to retain and how much of the new candidate activation $\widetilde{h_t}$ to consider. It is computed in Equation 1.

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{1}$$

where W_z and U_z are weight matrices associated with the update gate, and σ represents the sigmoid activation function.

The reset gate r_t determines how much of the previous hidden state to forget and how much of the new candidate activation to update the hidden state. It is calculated in Equation 2.

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{2}$$

where W_r and U_r are weight matrices associated with the reset gate.

The candidate activation $\widetilde{h_t}$ is a new proposal for the hidden state, combining information from the input x_t and the reset gate r_t. It is calculated in Equation 3.

$$\widetilde{h_t} = \tanh\big(W_h x_t + U_h (r_t \odot h_{t-1})\big) \tag{3}$$

where W_h and U_h are weight matrices associated with the candidate activation, $\odot$ denotes element-wise multiplication, and tanh represents the hyperbolic tangent activation function.

The updated hidden state h_t is computed by combining the previous hidden state h_{t-1} and the candidate's activation $\widetilde{h_t}$ using the update gate z_t, as shown in Equation 4.

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h_t} \tag{4}$$

These equations allow the GRU to selectively update and control the flow of information, making it capable of capturing both short-term and long-term dependencies in sequential data.

2.2. Stacked GRU

Assuming we have n GRU units in each layer, Equations 5-7 for the Stacked GRU model can be summarized as follows:

First GRU layer:

$$H_1 = \mathrm{GRU}(X) = \mathrm{GRU}(X; W_1, U_1) \tag{5}$$

Second GRU layer:

$$H_2 = \mathrm{GRU}(D_1) = \mathrm{GRU}(D_1; W_2, U_2) \tag{6}$$

Third GRU layer:

$$Y = \mathrm{GRU}(D_2) = \mathrm{GRU}(D_2; W_3, U_3) \tag{7}$$

where X is the input sequence, D_1 and D_2 represent the outputs of the dropout layers, H_1, H_2, and Y are the hidden states of the GRU layers, and W_i and U_i represent the weight matrices associated with the i-th GRU layer.

The stacked GRU architecture can effectively capture intricate patterns and long-term dependencies within the input sequence by incorporating multiple layers of GRU units. This enhanced architecture can potentially improve the accuracy and performance of stock price prediction models.

2.3. Dataset

This research employs a dataset from the Denpasar Weather Data, accessible to anyone on Kaggle. The data entails weather details from Denpasar, Indonesia, collected every hour from 1990 till the beginning of 2020, offering a substantial 20-year period of weather data.

The subset of the data transferred into this study includes 264,924 instances. Each instance incorporates four fundamental features - temperature, pressure, humidity, and wind speed. Such features are primary determinants in comprehending and forecasting weather conditions, which explains their common usage in weather prediction models. Their inclusion facilitates an in-depth study of their influence on the precision of weather predictions.

Further, the substantial quantity of the dataset and its hourly updates offer a solid backbone for creating and assessing weather prediction models. The goal of this dataset is to utilize the data within the Denpasar Weather Data to construct and gauge the effectiveness of stacked GRU models in accurate weather forecasting.

Refer to Figures 1-4 for a display of the weather pattern during the most recent 1000 hours.

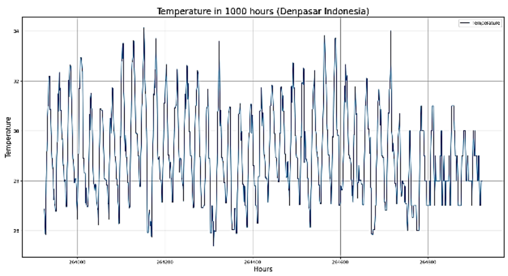

Figure 1. Temperature

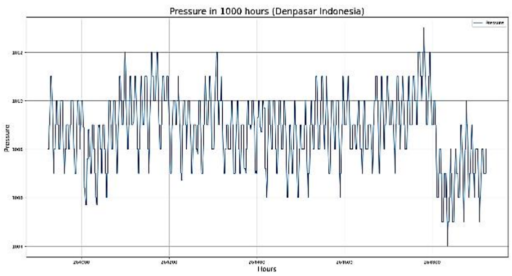

Figure 2. Pressure

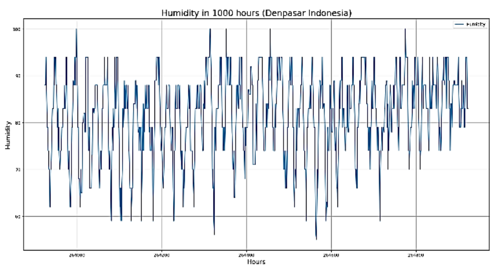

Figure 3. Humidity

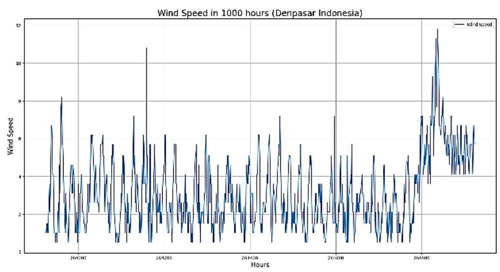

Figure 4. Wind speed

2.4. Data Preprocessing

Data processing is essential for preparing Denpasar Weather Data to be used in weather prediction models. Several crucial steps are taken to ensure the data is suitable and high-quality for analysis. The initial step is to manage missing values. This research removes instances containing missing values, preserving the dataset's integrity and ensuring accurate modeling and analysis.

After addressing missing values, relevant features for weather prediction are chosen. This study focuses on four significant features: temperature, pressure, humidity, and wind speed. These features significantly impact weather patterns and are vital for precise weather prediction modeling. Normalization is applied to the chosen features to guarantee compatibility in the modeling process. The Max-Min scaler is used for normalization, converting each feature's values into a common range between 0 and 1. Equation 8 shows the normalization equation.

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (8)$$

Here, X represents the original feature values, $X_{\min}$ denotes the minimum value of the feature, and $X_{\max}$ indicates the maximum value of the feature. Feature value normalization prevents biases caused by variations in scale or magnitude, enabling fair comparisons and accurate modeling.

By executing these data processing steps – managing missing values, choosing pertinent features, and applying the Max-Min scaler for data normalization – Denpasar Weather Data is primed for further analysis and modeling. This allows for accurate weather prediction using stacked GRU models.

2.5. Data Splitting

The train-test split method is essential for assessing the performance and generalizability of a weather prediction model. This research creates a balanced division between training and evaluation by splitting the Denpasar Weather Data into three subsets: training, validation, and test.

The dataset comprises 264,924 hourly data points, covering January 1, 1990, to January 7, 2020. The training set is formed by allocating 80% of the data, or 211,859 instances, for model training. This extensive training set helps the model learn from weather patterns across 20 years, fostering robust learning and capturing long-term dependencies in the data.

The validation set comprises 20% of the data, or 52,965 instances. This set acts as an intermediate step between training and final testing and helps evaluate the model's performance during training. The validation set detects potential overfitting or underfitting issues and optimizes hyperparameters and model architecture.

A separate test set of 240 instances is used for the final evaluation. This small test set gives an unbiased and independent measure of the model's prediction ability on unseen data. Its limited size allows fast evaluation and accurately reflects overall performance.

By partitioning the data into training, validation, and test sets as described, the method ensures that the model is trained on a large portion of the data, validated for performance tuning, and evaluated on unseen data for generalizability assessment. This approach provides a balanced evaluation and allows for a thorough examination of the model's accuracy and performance in weather prediction tasks.

2.6. Model Training Process

The model training process includes key aspects: selecting hyperparameters, defining the model architecture, choosing an optimization algorithm, and determining a suitable loss function.

1. Hyperparameters: This model has two main hyperparameters: the number of time steps (n_steps) and the number of features (n_features). In the code, n_steps is set to 240, meaning the model uses data from the past 240 hours to predict the next hour's weather. Since the dataset has a single feature (temperature, pressure, humidity, or wind speed), n_features is 1.

2. Model Architecture: The model has stacked GRU layers. The first GRU layer has 150 units and uses the 'relu' activation function. It returns sequences because additional GRU layers follow it. The second GRU layer also consists of 150 units and returns sequences. The final GRU layer has 150 units but does not return sequences. The model includes a Dense layer with one unit for the final output. This structure allows the model to learn complex patterns and dependencies in the input data.

3. Optimization Algorithm: The chosen optimization algorithm is 'adam', which stands for Adaptive Moment Estimation. Adam is widely used for training deep learning models efficiently. It merges

adaptive learning rates with momentum for faster convergence during training. Adam optimizer adjusts the model's weights and biases to minimize the loss function.

4. Loss Function: Mean Squared Error (MSE) is the selected loss function for this model. MSE is often used in regression tasks, such as weather prediction. It computes the average squared difference between predicted and actual values. By reducing the MSE loss, the model aims to decrease prediction errors and enhance the accuracy of weather forecasts.

The model is trained for ten epochs using the fixed architecture, optimizer, and loss function. It learns underlying patterns and relationships in the data to accurately predict the next hour's weather, utilizing the previous 240 hours' historical weather data.

2.7. Evaluation Metrics

The model's performance is assessed using common metrics such as RMSE, MAE, MAPE, and R2. Each metric offers insights into the model's prediction accuracy and fit. Equations 9-12 illustrate the calculations for these evaluation metrics.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100\% \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{12}$$

In these equations, n is the number of data points, y_i signifies the actual values, $\widehat{y_i}$ stands for the predicted values, and $\bar{y}$ represents the mean of the actual values.

These metrics deliver quantitative measures to evaluate the model's accuracy and performance in weather prediction. Lower RMSE and MAE values suggest better accuracy, and higher R2 values indicate an improved fit between predicted and actual values. MAPE shows the percentage difference between the predicted and actual values. By examining the model with these metrics, researchers can thoroughly determine its effectiveness in weather prediction tasks.

3. RESULTS AND ANALYSIS

3.1. Feature Correlation

As shown in Figure 5, here is the analysis of the correlation table between weather features of temperature, pressure, humidity, and wind speed:

1. Temperature and Pressure: A correlation coefficient of -0.399557 suggests a moderate negative correlation, which indicates that as temperature increases, the pressure typically decreases.
2. Temperature and Humidity: A correlation coefficient of -0.679107 suggests a significant negative correlation. Here, an increase in temperature tends to be associated with a decrease in humidity.
3. Temperature and Wind Speed: The correlation coefficient for these features is 0.205148, a low positive correlation. This indicates that temperature and wind speed only weakly increase together.
4. Pressure and Humidity: The correlation here is 0.058880, almost negligible, meaning changes in pressure do not regularly coincide with changes in humidity.
5. Pressure and Wind Speed: A correlation coefficient of 0.001580 suggests no significant relationship between pressure and wind speed.
6. Humidity and Wind Speed: The coefficient of -0.350623 signifies a moderate negative correlation. An increase in humidity implies a decrease in wind speed and vice versa.

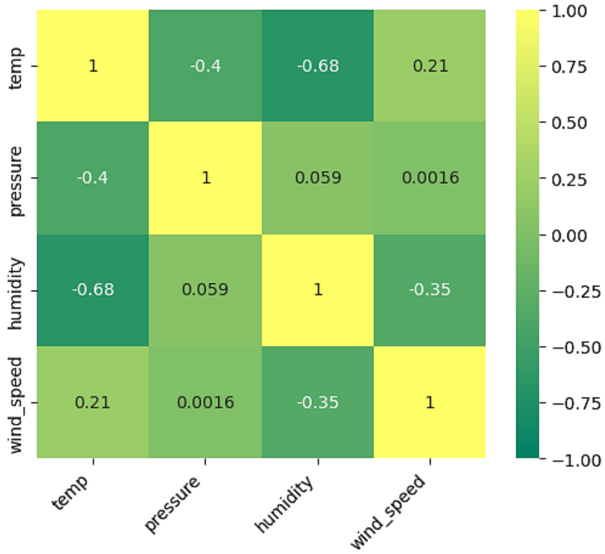

Figure 5. Feature Correlation

Correlation is a measure of association, not causation. It merely indicates the degree to which two variables move in relation to each other.

3.2. Performance Metrics

This study utilized the Stacked GRU model to predict the weather for the next 240 hours. The model's performance was evaluated using key metrics, including RMSE, MAE, MAPE, and R2, as shown in Table 1. These metrics provide insights into the accuracy and reliability of the predictions. The results obtained from the model demonstrated promising performance, with low values of RMSE and MAE, indicating small errors between the actual and predicted values. The MAPE values showed a reasonable percentage of error in the predictions, while the R2 values indicated a high level of variance explained by the model.

Table 1. Performance analysis of the Stacked GRU

Feature	RMSE	MAE	MAPE	R2
Temperature	0.03263	0.02770	0.04828	0.90975
Pressure	0.02950	0.02448	0.06125	0.94575
Humidity	0.04548	0.03554	0.05195	0.87628
Wind Speed	0.03782	0.02954	0.19125	0.94959
Average	0.03635	0.02932	0.08818	0.92034

The figures were generated to illustrate the performance visually, showcasing the actual and predicted weather for the next 240 hours, as shown in Figures 6-9. This graphical representation compares the predicted (red line) and observed (blue line) weather, highlighting any significant trends or deviations. The figure provides a visual confirmation of the model's ability to capture the general patterns and movements in the weather, further supporting the effectiveness of the Stacked GRU model in predicting future weather trends.

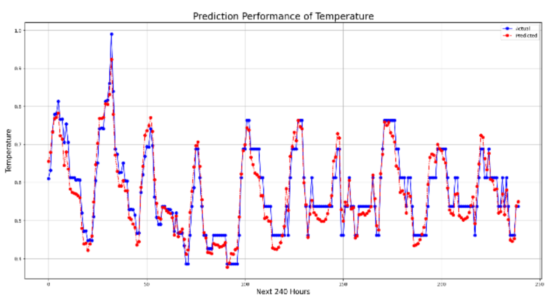

Figure 6. Temperature performance

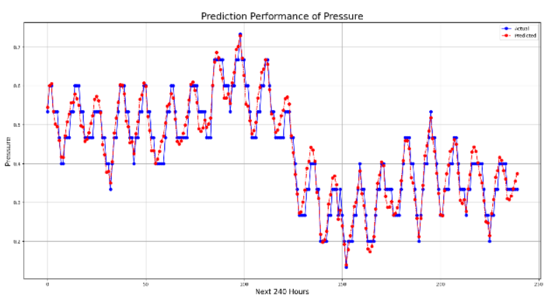

Figure 7. Pressure performance

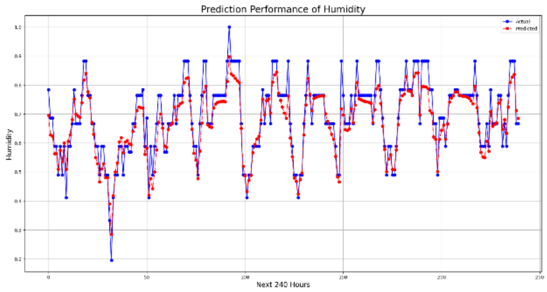

Figure 8. Humidity performance

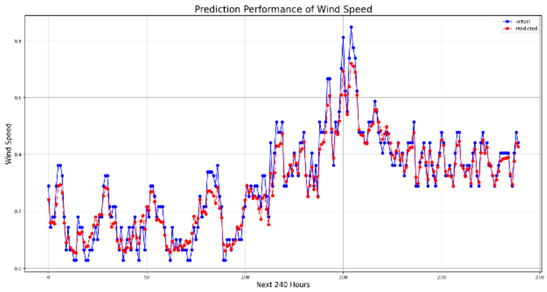

Figure 9. Wind speed performance

3.3. Performance Analysis

The Stacked GRU model's performance in predicting weather features can be analyzed using RMSE, MAE, MAPE, and R2 metrics. Each feature has its characteristics and challenges, as discussed below. The model shows a relatively low error in predicting temperature, as evidenced by the low RMSE, MAE, and MAPE values. With an R2 score of 0.90975, the model has strong predictive power for temperature. This suggests that the model captures most of the variance in temperature data and can accurately forecast temperature changes.

The pressure prediction performance is better than that of temperature, with lower RMSE, MAE, and MAPE values. The R2 score of 0.94575 indicates high reliability in predicting pressure. The model excels in understanding pressure data's underlying patterns and complexities, making it suitable for predicting pressure changes.

Humidity prediction has a higher RMSE, MAE, and MAPE than temperature and pressure, indicating that the model has more difficulty accurately predicting humidity. The R2 score, at 0.87628, is the lowest among the features, suggesting that the model has room for improvement in capturing the variance in humidity data. The model's performance in humidity forecasting may be affected by factors such as the complex interactions between temperature, pressure, and other weather phenomena.

Wind speed prediction has a lower RMSE and MAE than humidity but exhibits a notably higher MAPE value, indicating a larger percentage difference between predicted and actual values. Despite the high MAPE, the model has an impressive R2 score of 0.94959, suggesting that it explains a significant portion of the variance in wind speed data. The high MAPE value may be attributed to wind speed's inherent unpredictability and volatility due to various factors, such as geographical locations and complex atmospheric processes.

The Stacked GRU model demonstrates reliable performance across temperature, pressure, humidity, and wind speed predictions. While improvements can be made, particularly in predicting humidity and wind speed, the model has proven its potential for accurate and efficient weather forecasting.

3.4. Strengths and Limitations

The Stacked GRU model's strengths and limitations can be assessed based on its performance in predicting different weather features such as temperature, pressure, humidity, and wind speed.

The Stacked GRU model employed in this research possesses several strengths, as follows:

1. Accurate Predictions: Overall, the model shows good accuracy in predicting temperature, pressure, and wind speed, as evidenced by low RMSE and MAE values.
2. High R2 scores: The model displays high R2 scores for temperature, pressure, and wind speed predictions, indicating that it can effectively capture variance in these features' data.
3. Use of Temporal Information: As a recurrent neural network variant, the GRU model is particularly suited for time series data, like weather patterns, where temporal dynamics are key.

Despite its strengths, the Stacked GRU model also has certain limitations that should be considered, as follows:

1. Difficulty with Humidity Prediction: The model seems to struggle more with predicting humidity. The error rates for humidity (RMSE, MAE, and MAPE) are higher than those for other features.
2. High MAPE values: Despite the lower RMSE and MAE values, the significantly high MAPE value for wind speed suggests that there can be large percentage differences between the predicted and actual values. This could mean that the model might sometimes give significantly off predictions.
3. Potential Overfitting: The model's high accuracy may also indicate overfitting, where it has learned the training data too well but may not perform well on new, unseen data.
4. Computational Expense: Stacked GRU can be computationally expensive and time-consuming to train due to the stacking of multiple layers of GRUs.

In summary, limitations exist while the Stacked GRU has demonstrated strengths, particularly where handling temporal dynamics is key. Considering these when deploying the model or interpreting its results will be essential.

3.4. Implications of The Findings

This research highlights the potential of the Stacked GRU model in improving weather prediction accuracy for various weather variables. The model's ability to accurately forecast temperature, pressure, humidity, and wind speed showcases its relevance and significance in critical sectors such as agriculture, transportation, and disaster preparedness. By leveraging deep learning techniques, this study contributes to advancements in weather forecasting, empowering decision-makers with precise information for informed planning, resource management, and risk mitigation. The findings underscore the importance of adopting sophisticated machine learning approaches to enhance weather prediction capabilities, facilitating more effective strategies for weather-related applications and improving overall societal resilience in changing weather patterns.

3.5. Practical Recommendations

For researchers and practitioners looking to leverage Stacked GRU for weather prediction, practical recommendations include: (1) Conducting comprehensive data preprocessing to ensure high data quality and consistency, addressing missing values, outliers, and scaling data appropriately. (2) Fine-tuning the model architecture and hyperparameters through systematic experimentation to optimize performance and avoid overfitting. (3) Using ensembles to combine predictions from multiple Stacked GRU models enhances robustness and prediction accuracy. (4) Exploring transfer learning techniques by pretraining the model on related weather datasets to leverage existing knowledge and improve performance on specific regions or weather patterns. (5) Focusing on model interpretability by incorporating attention mechanisms or other explainable AI techniques to gain insights into the model's predictions and foster trust in the results. By following these recommendations, researchers and practitioners can effectively harness the power of Stacked GRU for improved weather prediction outcomes and advance the application of deep learning in weather forecasting.

4. CONCLUSION

The Stacked Gated Recurrent Unit (GRU) model demonstrates notable capability in weather forecasting. It has achieved high accuracy in predicting temperature, pressure, humidity, and wind speed, verified by low RMSE and MAE values, alongside high R2 scores. These results suggest the model's predictive strength, displaying its aptitude for handling time series weather data. However, the model faces challenges in predicting humidity, as seen from higher error rates in this area, and there are concerns about significant discrepancies in percentage errors within wind speed forecasting. These areas represent potential boundaries of the model that need to be addressed in future research.

As we look ahead, these results provide a foundation for future study. The possibility of improving humidity prediction and addressing the discrepancy in wind speed forecasting offers avenues for further development. Potential overfitting and computational expense also present opportunities for exploration and optimization to enhance model efficiency and scalability. Optimized versions of this model bear potential for applications beyond weather forecasting. Numerous fields involving time-series data predictions, including stock market forecasting, energy demand prediction, and traffic flow prediction, may benefit from similar advancements. Ultimately, this research has highlighted the power of the Stacked GRU model for weather forecasting but also established areas that need refinement for optimal performance. The future holds promising opportunities for both the enhancement of this model and the extension of similar models to other applications.

REFERENCES

[1] G. Zuma-Netshiukhwi, O. M. Hlazo, and S. A. Motholo, “Evaluating the Effect of Climate Variability on Zea Mays Productivity Over Glen Research Station: South Africa,” *European Journal of Agriculture and Food Sciences*, 2021, doi: 10.24018/ejfood.2021.3.3.313.

[2] R. Varahasamy*, S. Meganathan, and D. Karthik, “Association Rule Mining for South West Monsoon Rainfall Prediction and Estimation Over Mumbai Station,” *International Journal of Recent Technology and Engineering*, 2019, doi: 10.35940/ijrte.c6809.098319.

[3] Z. Zhang and Y. Dong, “Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data,” *Complexity*, 2020, doi: 10.1155/2020/3536572.

[4] Y. Xie, J. Shi, S. Fan, M. Chen, Y. Dou, and D. Ji, “Impact of Radiance Data Assimilation on the Prediction of Heavy Rainfall in RMAPS: A Case Study,” *Remote Sensing*, 2018, doi: 10.3390/rs10091380.

[5] T. Ushiyama, T. Sayama, and Y. Iwami, "Development of a Flood Forecasting System Using Regional Ensemble Prediction – Application to the Kinugawa Flood in 2015," *Journal of Jsce*, 2019, doi: 10.2208/journalofjsce.7.1_124.

[6] T. Röösli, C. Appenzeller, and D. N. Bresch, "Towards Operational Impact Forecasting of Building Damage From Winter Windstorms in Switzerland," *Meteorological Applications*, 2021, doi: 10.1002/met.2035.

[7] J. Jing, Q. J. Li, X. H. Ding, N. X. Sun, R. Tang, and Y. L. Cai, "Aenn: A Generative Adversarial Neural Network for Weather Radar Echo Extrapolation," *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, 2019, doi: 10.5194/isprs-archives-xlii-3-w9-89-2019.

[8] T. Andrasto, N. Musaropah, N. Haryono, T. Joko, and N. Kardoyo, "Simulation and Design of Smart Clothesline Using Fuzzy for Weather Forecast," *Iop Conference Series Earth and Environmental Science*, 2022, doi: 10.1088/1755-1315/969/1/012058.

[9] T. P. Agyekum, P. Antwi-Agyei, and A. J. Dougill, "The contribution of weather forecast information to agriculture, water, and energy sectors in East and West Africa: A systematic review," *Frontiers in Environmental Science*, vol. 10, 2022, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2022.935696

[10] H. Sangrody, M. Sarailoo, N. Zhou, N. Tran, M. Motalleb, and E. Foruzan, "Weather Forecasting Error in Solar Energy Forecasting," *Iet Renewable Power Generation*, 2017, doi: 10.1049/iet-rpg.2016.1043.

[11] L. van Galen, O. Hartogensis, I. Benedict, and G.-J. Steeneveld, "Teaching a Weather Forecasting Class in the 2020s," *Bulletin of the American Meteorological Society*, 2022, doi: 10.1175/bams-d-20-0107.1.

[12] I. Sideris, L. Foresti, D. Nerini, and U. Germann, "NowPrecip: Localized Precipitation Nowcasting in the Complex Terrain of Switzerland," *Quarterly Journal of the Royal Meteorological Society*, 2020, doi: 10.1002/qj.3766.

[13] R. Ramadhan *et al.*, "Evaluation of GPM IMERG Performance Using Gauge Data Over Indonesian Maritime Continent at Different Time Scales," *Remote Sensing*, 2022, doi: 10.3390/rs14051172.

[14] Y. Wang *et al.*, "An Improved ENSO Ensemble Forecasting Strategy Based on Multiple Coupled Model Initialization Parameters," *Journal of Advances in Modeling Earth Systems*, 2019, doi: 10.1029/2019ms001620.

[15] A. S. Santra and J. Lin, "Integrating Long Short-Term Memory and Genetic Algorithm for Short-Term Load Forecasting," *Energies*, 2019, doi: 10.3390/en12112040.

[16] M. C. Lopes and R. Lamberts, "Development of a Metamodel to Predict Cooling Energy Consumption of HVAC Systems in Office Buildings in Different Climates," *Sustainability*, 2018, doi: 10.3390/su10124718.

[17] M. J. Pavolonis, J. Sieglaff, and J. L. Cintineo, "Automated Detection of Explosive Volcanic Eruptions Using Satellite-Derived Cloud Vertical Growth Rates," *Earth and Space Science*, 2018, doi: 10.1029/2018ea000410.

[18] H. Sijie, L. Xinyan, T. Laurie, C. A Benjamin, D. Timothy, and B. Arindam, "Learning and Dynamical Models for Sub-Seasonal Climate Forecasting: Comparison and Collaboration," 2021, doi: 10.48550/arxiv.2110.05196.

[19] Y. Geng *et al.*, "A Deep Learning Framework for Lightning Forecasting With Multi-source Spatiotemporal Data," *Quarterly Journal of the Royal Meteorological Society*, 2021, doi: 10.1002/qj.4167.

[20] K. Ingenloff and A. C. Peterson, "Incorporating Time Into the Traditional Correlational Distributional Modelling Framework: A Proof-of-concept Using the Wood Thrush Hylocichla Mustelina," *Methods in Ecology and Evolution*, 2020, doi: 10.1111/2041-210x.13523.

[21] J. Xia *et al.*, "Machine Learning-Based Weather Support for the 2022 Winter Olympics," *Advances in Atmospheric Sciences*, 2020, doi: 10.1007/s00376-020-0043-5.

[22] R. Meenal, P. A. Michael, and D. Pamela, "Weather Prediction Using Random Forest Machine Learning Model," *Indonesian Journal of Electrical Engineering and Computer Science*, 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.

[23] C.-Y. Chang, C.-W. Lu, and C.-J. Wang, "A Multi-Step-Ahead Markov Conditional Forward Model With Cube Perturbations for Extreme Weather Forecasting," *Proceedings of the Aaai Conference on Artificial Intelligence*, 2021, doi: 10.1609/aaai.v35i8.16856.

[24] M. Abdulraheem, J. B. Awotunde, E. A. Adeniyi, I. D. Oladipo, and S. O. Adekola, "Weather Prediction Performance Evaluation on Selected Machine Learning Algorithms," *Iaes International Journal of Artificial Intelligence (Ij-Ai)*, 2022, doi: 10.11591/ijai.v11.i4.pp1535-1544.

[25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," 2014, doi: 10.48550/arxiv.1412.3555.

[26] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research*, 2019, doi: 10.2478/jaiscr-2019-0006.

[27] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "Improving Speech Recognition by Revising Gated Recurrent Units," 2017, doi: 10.21437/interspeech.2017-775.

[28] G. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, "Minimal Gated Unit for Recurrent Neural Networks," *International Journal of Automation and Computing*, 2016, doi: 10.1007/s11633-016-1006-2.

[29] A. Zabardast, E. G. Tamer, Y. A. Son, and A. Yılmaz, "An Automated Framework for Evaluation of Deep Learning Models for Splice Site Predictions," *Scientific Reports*, 2023, doi: 10.1038/s41598-023-34795-4.

[30] N. C. S. Alfredo and D. A. Adytia, "Time Series Forecasting of Significant Wave Height Using GRU, CNN-GRU, and LSTM," *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, 2022, doi: 10.29207/resti.v6i5.4160.

[31] Z. Zeng, X. Tang, Y. Liu, Z. He, and X. Gong, "Interpretable Recurrent Neural Network Models for Dynamic Prediction of the Extubation Failure Risk in Patients With Invasive Mechanical Ventilation in the Intensive Care Unit," *Biodata Mining*, 2022, doi: 10.1186/s13040-022-00309-7.

[32] A. Gasparin, S. Lukovic, and C. Alippi, "Deep Learning for Time Series Forecasting: The Electric Load Case," *Caai Transactions on Intelligence Technology*, 2021, doi: 10.1049/cit2.12060.

[33] S.-W. Jung, J. Moon, S. W. Park, and E. Hwang, "An Attention-Based Multilayer GRU Model for Multistep-Ahead Short-Term Load Forecasting," *Sensors*, 2021, doi: 10.3390/s21051639.
[34] J. Wojtkiewicz, M. Hosseini, R. Gottumukkala, and T. L. Chambers, "Hour-Ahead Solar Irradiance Forecasting Using Multivariate Gated Recurrent Units," *Energies*, 2019, doi: 10.3390/en12214055.
[35] X.-B. Jin, X. H. Yu, X. Wang, Y. Bai, T. Su, and J.-L. Kong, "Deep Learning Predictor for Sustainable Precision Agriculture Based on Internet of Things System," *Sustainability*, 2020, doi: 10.3390/su12041433.
[36] J. Zang *et al.*, "Prediction Model of Carbon Dioxide Concentration in Pig House Based on Deep Learning," *Atmosphere*, 2022, doi: 10.3390/atmos13071130.
[37] N. Li, H. Sheng, P. Wang, Y. Jia, Z. Yang, and Z. Jin, "Modeling Categorized Truck Arrivals at Ports: Big Data for Traffic Prediction," *Ieee Transactions on Intelligent Transportation Systems*, 2023, doi: 10.1109/tits.2022.3219882.
[38] B. Yang, T. Xiao, and L. Wang, "Using Complementary Ensemble Empirical Mode Decomposition and Gated Recurrent Unit to Predict Landslide Displacements in Dam Reservoir," *Sensors*, 2022, doi: 10.3390/s22041320.
[39] F. R. Alharbi and D. Csala, "Short-Term Wind Speed and Temperature Forecasting Model Based on Gated Recurrent Unit Neural Networks," 2021, doi: 10.1109/gpecom52585.2021.9587479.
[40] Y. Zhao and Y. Ban, "Early Detection of Wildfires With GOES-R Time Series and Deep GRU-Network," 2022, doi: 10.20944/preprints202207.0265.v1.
[41] S. Boubaker, M. Benghanem, A. Lefza, O. Kahouli, and L. Kolsi, "Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia," *Ieee Access*, 2021, doi: 10.1109/access.2021.3062205.
[42] S. Roy, I. Kiral-Kornek, and S. Harrer, "ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification," 2019, doi: 10.1007/978-3-030-21642-9_8.
[43] A. Caniago, W. Kaswidjanti, and J. Juwairiah, "Recurrent Neural Network With Gate Recurrent Unit for Stock Price Prediction," *Telematika*, 2021, doi: 10.31315/telematika.v18i3.6650.
[44] R. K. Pandit, D. Astolfi, A. M. Tang, and D. Infield, "Sequential Data-Driven Long-Term Weather Forecasting Models' Performance Comparison for Improving Offshore Operation and Maintenance Operations," *Energies*, 2022, doi: 10.3390/en15197233.
[45] S. Scher and G. Messori, "Predicting Weather Forecast Uncertainty With Machine Learning," *Quarterly Journal of the Royal Meteorological Society*, 2018, doi: 10.1002/qj.3410.
[46] N. Aslam, F. Rustam, E. A. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," *Ieee Access*, 2022, doi: 10.1109/access.2022.3165621.
[47] C. Zeng, C. Ma, K. Wang, and Z. Cui, "Parking Occupancy Prediction Method Based on Multi Factors and Stacked GRU-LSTM," *Ieee Access*, 2022, doi: 10.1109/access.2022.3171330.
[48] J. K. Mandal, T. Kim, and S. K. Bandyopadhyay, "Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN," *Applied Computer Systems*, 2020, doi: 10.2478/acss-2020-0018.

BIBLIOGRAPHY OF AUTHORS

Mohammad Diqi is a researcher affiliated with the Department of Informatics at Universitas Respati Yogyakarta, Indonesia. His research interests lie in the fields of data science and machine learning. He has contributed to various academic works and publications, exploring cutting-edge advancements and innovative approaches in these domains. With a focus on leveraging data-driven methodologies, his bibliography encompasses research papers, conference presentations, and contributions to the field that have enriched the scientific community's understanding and application of data science and machine learning techniques.

Ahmad Wakhid is a researcher affiliated with the School of Computer Science and Technology at Harbin University of Science and Technology, China. His research interests span diverse areas, including Internet of Things (IoT), Electronic Development, and Project Management. Throughout his academic journey, Wakhid has made significant contributions to the fields of IoT and electronic development, exploring the integration of smart devices, sensors, and communication technologies to create interconnected systems with real-world applications. Additionally, his expertise in project management has enabled him to effectively lead and execute research initiatives, ensuring the successful implementation of complex projects in the realm of technology and innovation.

I Wayan Ordiyasa is a researcher affiliated with the Department of Informatics at Universitas Respati Yogyakarta, Indonesia. His research is focused on the exciting domains of the Internet of Things (IoT) and Metaverse. He explores integrating IoT technologies with virtual environments, aiming to create immersive and interactive experiences within the Metaverse. Through his academic pursuits, Wayan investigates novel ways to connect IoT devices and sensors to virtual worlds, enabling seamless interactions and data exchange between the physical and virtual realms. His contributions in these fields aim to push the boundaries of technological innovation and unlock the full potential of the IoT and Metaverse, offering transformative applications and possibilities for the future.

Nurhadi Wijaya is a researcher affiliated with the Department of Informatics at Universitas Respati Yogyakarta, Indonesia. His research focuses on the intersection of the Internet of Things (IoT) and machine learning. Through his academic works and publications, he delves into integrating smart devices and IoT technologies with machine learning algorithms, aiming to enhance the capabilities of IoT systems and enable intelligent decision-making processes. His contributions in this field have provided valuable insights and innovative solutions, fostering advancements in IoT and machine learning for practical applications.

Marselina Endah Hiswati is a researcher affiliated with the Department of Informatics at Universitas Respati Yogyakarta, Indonesia. Her research revolves around machine learning and computational thinking. With a passion for exploring the principles and methodologies of machine learning, she seeks to develop novel algorithms and models that can enhance the understanding and prediction of complex patterns within data. Additionally, she investigates the application of computational thinking in problem-solving and decision-making processes, exploring how this cognitive approach can empower individuals to tackle real-world challenges effectively. Through her academic contributions, Marselina aims to advance the field of machine learning and promote the adoption of computational thinking techniques in various domains for societal benefits.