Scientific Journal of Informatics
Vol. 10, No. 4, Nov 2023

p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Exploring Long Short-Term Memory and Gated Recurrent Unit Networks for Emotion Classification from Electroencephalography Signals

Dian Palupi Rini[1*], Winda Kurnia Sari[2], Novi Yusliani[3], Deris Stiawan[4] Aspirani Utari[5]
[1,3]Department of Informatics Engineering, Faculty of Computer Science,
Universitas Sriwijaya, Indonesia
[2,5]Department of Information System, Faculty of Computer Science,
Universitas Sriwijaya, Indonesia
[4]Department of Computer System, Faculty of Computer Science,
Universitas Sriwijaya, Indonesia

Abstract. This study delves into comparing LSTM and GRU, two recurrent neural network (RNN) models, for classifying emotion data through electroencephalography (EEG) signals. Both models adeptly handle sequential data challenges, showcasing their unique strengths. In EEG emotion dataset experiments, LSTM demonstrated superior performance in emotion classification compared to GRU, despite GRU's quicker training processes. Evaluation metrics encompassing accuracy, recall, F1-score, and area under the curve (AUC) underscored LSTM's dominance, which was particularly evident in the ROC curve analysis. This research sheds light on the nuanced capabilities of these RNN models, offering valuable insights into their efficacy in emotion classification tasks based on EEG data. The study explores parameters, such as the number of layers, neurons, and the utilization of dropout, providing a detailed analysis of their impact on emotion recognition accuracy.
Purpose: The proposed model in this study is the result of optimizing LSTM and GRU networks through careful parameter tuning to find the best model for classifying EEG emotion data. The experimental results indicate that the LSTM model can achieve an accuracy level of up to 100%.
Methods: To improve the accuracy of the LSTM and GRU methods in this research, hyperparameter tuning techniques were applied, such as adding layers, dense layers, flatten layers, selecting the number of neurons, and introducing dropout to mitigate the risk of overfitting. The goal was to find the best model for both methods.
Results: The proposed model in this study is capable of classifying EEG emotion data very effectively. The experimental results demonstrate that the LSTM model achieves a maximum accuracy of 100%, while the GRU model achieves a highest accuracy of approximately 98%.
Novelty: The novelty of this research lies in the optimization of hyperparameters for both LSTM and GRU methods, leading to the development of novel architectures capable of effectively classifying EEG emotion data.

Keywords: Long short-term memory; Gated recurrent unit; Emotion classification; EEG signals.
Received September 2023 / **Revised** October 2023 / **Accepted** November 2023

INTRODUCTION

The expression of human emotions is a complex manifestation of various psychological and physiological aspects, often related to subjective feelings, temperaments, personalities, motivational inclinations, behavioral responses, and physiological stimuli [1], [2]. Behavioral and physiological signals have been investigated to recognize human emotions. Commonly used behavioral signals include speech, facial expressions, and hand and body movements [3]. Compared to behavioral signals that are easily hidden in emotion recognition, physiological measurements are more reliable for recognizing human emotions [4].

Model recognition and classification are important components in recognizing emotions based on EEG signals. Their main task is to identify EEG models that correspond to various emotional states by extracting different types of EEG features and then classifying the features of untrained EEG signals. The selection of the optimal classification model plays a very important role in emotion recognition because it can effectively improve the accuracy of emotion classification. Currently, common methods for recognizing and classifying emotions based on EEG signals involve machine learning and deep network learning. Along

*Dian Palupi Rini.
Email addresses: dprini@unsri.ac.id (Rini), windakurniasari@unsri.ac.id (Sari)
DOI: 10.15294/sji.v10i4.47734

with the rapid development of artificial intelligence, machine learning is emerging as a promising classification method [5], [6], [7].

Research on classification models for emotion recognition based on EEG signals using deep learning has been widely conducted, such as convolutional neural network (CNN) [8], [9], [10], recurrent neural network (RNN) [11], [6], [12], long short-term memory (LSTM) [13], [14], gated recurrent unit (GRU) [15], [16], and deep neural network (DNN) [17], [18], [19]. Deep learning methods are capable of automated learning from start to finish in EEG signal preprocessing, feature extraction, and classification. Emotion recognition based on deep learning algorithms on EEG signals has higher feature dimensions and shorter recognition times compared to manually designed features, so it can be a better support in medical diagnosis.

This study delves into the comparative performance analysis of two recurrent neural network architectures, long short-term memory (LSTM) and gated recurrent unit (GRU), within the realm of EEG-based emotion classification. Our comprehensive investigation scrutinizes their aptitude for handling intricate temporal data extracted from EEG signals, which chronicle brain activity over time. This research illuminates the nuanced disparities in performance between LSTM and GRU, providing invaluable insights into their roles in emotion classification tasks and serving as a guiding beacon for employing these recurrent neural network architectures in the intricate realm of temporal data processing. Notably, the study underscores the existing research gap, urging for a deeper understanding of the intricate neural patterns and emphasizing the substantial contribution this study makes in bridging this gap. The findings of this research are poised to elevate the comprehension and application of EEG-based emotion recognition techniques, significantly impacting the domains of medicine and cognitive science.

METHODS

Proposed Methodology

Generally, the steps applied in this research methodology involve the use of a structured framework to guide each stage. The research framework, as depicted in Figure 1, begins with a literature review stage encompassing the evaluation of studies conducted within the past 1 to 5 years. Moving on to the data preparation phase, the study utilized an emotion EEG dataset comprising more than 2,100 data samples. Subsequently, the preprocessing stage involved transforming categories or data labels into numerical representations. The classification process employed the recurrent neural network (RNN) architecture, specifically LSTM and GRU, and encompassed three key stages: training, validation, and testing. Additionally, it is noteworthy that, apart from the proposed methods, there exists a research flowchart illustrated in Figure 2, which provides a visual representation of the research process.

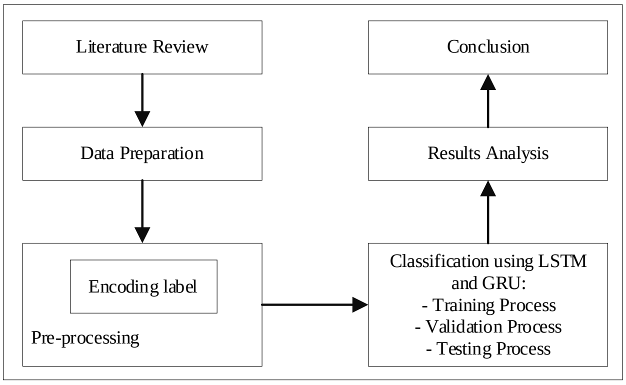

Figure 1. Proposed methodology

Data Preparation

The data for this research was collected from two subjects, a male and a female aged 20–22 years old. Data collection was conducted using four dry extracranial electrodes via a commercially available MUSE EEG headband, which made it possible to record EEG activity at the TP9, AF7, AF8, and TP10 location points [20], [21]. This dataset is further described in Table 1, with the following specifications: 60 seconds of data were recorded from 6 film clips, resulting in a total of 12 minutes of brain activity, including neutral emotional data. Neutral data collection was performed without stimuli before the emotional data collection. Although the data distribution among classes was uneven due to variations in the subjects' emotional responses, this variability provided crucial diversity for emotion analysis. This variability offers significant insights, enabling researchers to interpret experimental results more comprehensively.

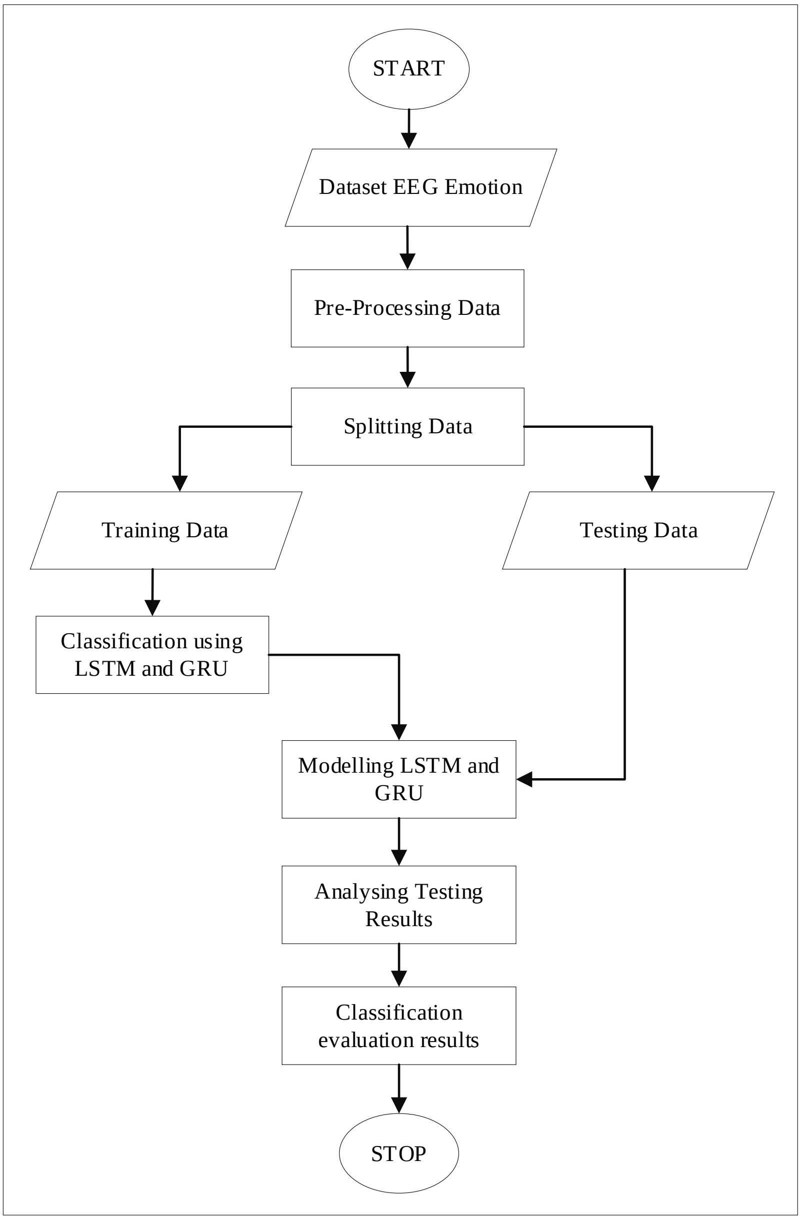

Figure 2. Research flow diagram

In the framework of the experiment, EEG datasets were used that focused on capturing brainwave patterns associated with various emotions and feelings. The dataset consists of 2,549 variables and 2,132 rows of data. Among these variables, 2,548 contain data in decimal format, while 1 other variable contains data in the form of strings that serve as labels.

After preparing the dataset, the next crucial step is to meticulously check for the presence of missing values, as indicated in the Python code in Figure 3. The importance of checking for missing values cannot be overlooked in the data analysis. The presence of missing values can affect the interpretation of the research results and the accuracy of the models used. Therefore, this step is fundamental to ensuring the reliability and precision of the data analysis process within the context of this research.

Table 1. Dataset Specifications

Data label	Number of data
Positive	708
Negative	708
Neutral	716
Total	2,132

```
# Checking for missing values
for col in df.columns:
    if(df[col].isnull().sum()>0):
        print(col)
```

Figure 3. Checking missing values

Preprocessing

In this study, label encoding was used to convert class variables that were originally in the form of string data into numeric values. The process of transforming labels or classes with pseudocode is shown in Figure 4. In the figure, three classes were originally in string form, originally in string form, namely: “Neutral” was changed to Class 0, “Positive” was changed to Class 1, and “Negative” was changed to Class 2.

```
BEGIN

FUNCTION label_encoder(category):
    IF category == "Neutral" THEN
        RETURN 0
    ELSE IF category == "Positive" THEN
        RETURN 1
    ELSE IF category == "Negative" THEN
        RETURN 2
    ELSE
        PRINT "Invalid category"
    END IF
END FUNCTION

// Example usage of label_encoder
input_category = "Positive"
label_encoded = label_encoder(input_category)
PRINT "Label encoded for category", input_category, "is", label_encoded

END
```

Figure 4. Label encoding

In the context of machine learning algorithms, categorical data cannot be directly processed by these algorithms. Therefore, categorical data must be converted into numerical form before they can be used in the analysis process [22]. This is important in this research because it focuses on sequence classification types that rely on deep learning methods, such as long short-term memory. Within this framework, converting categorical data into numerical representations is a necessary step to enable the algorithm to recognize and process sequential patterns related to the classification goal [23], [24].

Long Short-Term Memory

The long short-term memory (LSTM) method stands out in this research due to its various advantages. First, LSTM can capture long-term temporal relationships in sequential data, especially EEG signals that record brain activity over time. This ability enables LSTM to identify complex patterns related to human emotional changes. Second, LSTM is equipped with smart mechanisms that allow it to overcome the problem of long-term information loss. These mechanisms, consisting of input gates, forget gates, and output gates, enable LSTM to store, forget, and utilize information wisely in the decision-making process

[25]. Third, LSTM can address the common issue of vanishing gradients in conventional recurrent neural networks. With this capability, LSTM can maintain gradient flow during training, ensuring effective and accurate learning, especially in data with very long sequences. With these combined strengths, LSTM has proven to be highly effective in tackling complex challenges in EEG-based emotion recognition and delivering superior results in emotion classification [14].

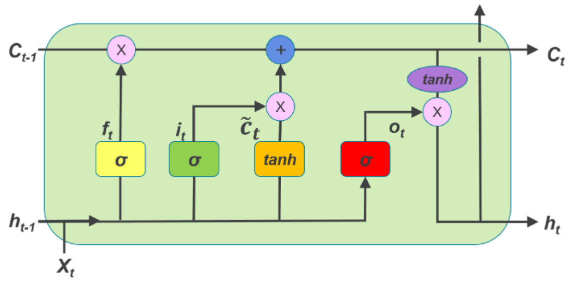

Figure 5. LSTM architecture

Below is the equation formula for all LSTM gates:

Input Gate:

$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i)$ (1)

$\check{C}_t = \tanh\,(W_C . [h_{t-1}, x_t] + b_C$ (2)

Input Gate:

$f_t = \sigma\big(W_f . [h_{t-1}, x_t + b_f\big)$ (3)

Input Gate:

$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ (4)

The process of making models in this study began with building input layers according to the form of exercise data. Then, three successive LSTM layers with neuron sizes of 256, 128, and 64 were implemented, all designed to produce sequential outputs. After that, a dense layer with 64 neurons and ReLU activation was used to continue extracting features from the LSTM output. To avoid overfitting, apply a dropout layer with a level of 0.5 after the dense layer. Next, another dense layer with 32 neurons and ReLU activation was added to further process the resulting features. In the end, an output layer with three neurons and softmax activation was used to generate classification predictions in a multi-class context. The complete model was assembled by combining the input and output layers using Keras' model class. These steps are designed to create LSTM models that effectively process sequential data and perform classification tasks with high accuracy.

Gated Recurrent Units

The GRU learning model has a more complex structure compared to conventional RNNs, such as long short-term memory (LSTM) [26], [27]. Such a structure allows the GRU to overcome some problems in EEG signal processing. One of the important impacts of GRU on EEG signal results is its ability to capture more complex temporal relationships in EEG data [28]. Because EEG signals are sequential data that record brain activity over time, temporal patterns are important in recognizing emotions and other cognitive processes. GRU can recognize and understand these patterns better, thus improving classification and emotion recognition skills.

GRU is a gating mechanism in RNN, similar to LSTM, which has a forget gate but with a smaller number of parameters because it does not include an output gate. Although LSTM is more powerful than GRU due to its ability to perform more complex calculations, GRU still has significant usefulness [29]. An overview of the GRU architecture can be seen in Figure 6.

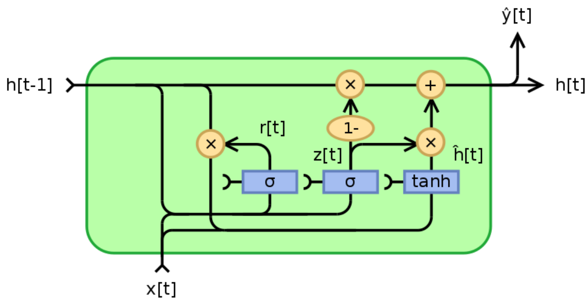

Figure 6. GRU architecture

To make comparisons and obtain the best model, the same model built using LSTM was constructed using GRU. This process started by defining the input layer according to the shape of the exercise data. Next, three successive GRU layers with neuron sizes of 256, 128, and 64 were implemented, all designed to produce sequential outputs. The data obtained from the GRU layers were then further processed using the flatten layer to convert it into a 1D format. A dense layer with 64 neurons and ReLU activation was used to continue the feature extraction. To reduce the risk of overfitting, apply a dropout layer with a level of 0.5 after the dense layer. Next, a dense layer with 32 neurons and ReLU activation was applied to further process the resulting features. Finally, an output layer with three neurons and softmax activation was used to generate classification predictions in a multi-class context. By performing similar steps to the LSTM model, these GRU models can be compared to obtain the best results.

Confusion Matrix

In machine learning, classifying data into more than two classes is referred to as multi-class classification. Performance metrics are crucial when evaluating and comparing different classification models or machine learning techniques. Various metrics prove useful for testing the proficiency of multi-class classifiers and are applied at different stages of development, such as comparing different models or analyzing a single model's behavior with adjusted parameters [30], [31].

Table 2. Confusion matrix class 0

Label 0		Actual		
		0	1	2
Prediction	0	TP_i	FP_i	FP_i
	1	FN_i	TN_i	
	2	FN_i		TN_i

Table 3. Confusion matrix class 1

Label 1		Actual		
		0	1	2
Prediction	0	TN_i	FN_i	
	1	FP_i	TP_i	FP_i
	2		FN_i	TN_i

Table 4. Confusion matrix class 2

Label 2		Actual		
		0	1	2
Prediction	0	TN_i		FN_i
	1		TN_i	FN_i
	2	FP_i	FP_i	TP_i

In this study, a multiclass confusion matrix with 3 classes was used, where calculating the accuracy, precision, recall, and F1-score values as shown in Equations 1–4 was done by calculating TPi, TNi, FPi, and FNi for each class. These values were then divided by the number of classes (l). The placement of TPi, TNi, FPi, and FNi in the confusion matrix was done for Class 0 in Table 2, Class 1 in Table 3, and Class 2 in Table 4.

Data Testing and Data Training Separation

The ratio of training, validation, and test data plays a critical role in shaping the outcomes in machine learning model training. If the proportion of training data is too low, the model may not learn patterns effectively and could potentially overfit the training data, limiting its ability to predict new data accurately. Conversely, if the proportion of training data is too high, the model might not have enough test data for evaluation, leading to a lack of generalization to unseen data. Proper allocation between training, validation, and test data is key to ensuring that machine learning models effectively learn patterns from training data, validate efficiently on validation data, and make accurate predictions on test data, ensuring robust and reliable model performance in real-world scenarios [32], [33]. The distribution of training and test data was carried out in five scenarios, as presented in Table 5.

Table 5. Data distribution

Validation	Data
First Validation	Data training 90% and data testing 10%
Second Validation	Data training 80% and data testing 20%
Third Validation	Data training 70% and data testing 30%
Fourth Validation	Data training 60% and data testing 40%
Fifth Validation	Data training 50% and data testing 50%

RESULTS AND DISCUSSIONS

In this study, a comparison was made between the long short-term memory (LSTM) model and the gated recurrent unit (GRU) model in emotion recognition based on EEG data. The comparison results show that, although both models use the same number of layers, the LSTM model has a slightly higher accuracy rate than the GRU model in various training and test data-sharing scenarios. These results show that, although the model architectures are similar, there are performance differences that can be attributed to the characteristics of each gating mechanism in LSTM and GRU.

The factor that may influence the suboptimal results of the GRU model is the interaction between the gate components present in the GRU architecture. Although GRU has fewer parameters than LSTM, the interaction between input gates, reset gates, and actual units in a GRU may have a different impact on the model’s ability to remember sequential information. In addition, the influence of the same number of layers on both models can provide insight into how GRU and LSTM architectures respond to the complexity of EEG data.

Although the GRU results were not better in this study, they suggested that the choice between LSTM and GRU should be based on the specific goals and characteristics of the emotion recognition task based on EEG data. The possibility of other factors influencing the results may also be explored further in future studies. Further discussion of the interpretation of the results and the practical implications of this comparison will provide a more in-depth look at the advantages and limitations of each model in the context of emotion recognition.

First Validation Result

In the first validation, 90% of the training data and 10% of the test data were divided into identical parameters and a number of layers in the LSTM and GRU models. The results of this validation are represented through the accuracy ROC graphs, which can be found in Figures 7 and 8. In addition, the development of the ROC loss curve can also be observed through visualization in Figures 9 and 10.

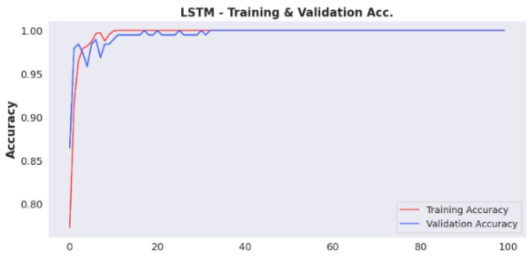

Figure 7. LSTM accuracy curve

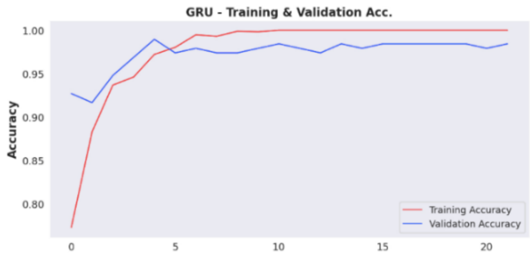

Figure 8. GRU accuracy curve

Figure 9. LSTM loss curve

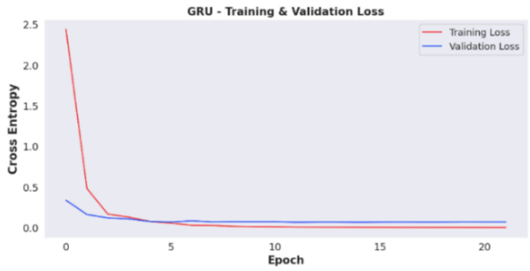

Figure 10. GRU loss curve

The results of the first validation are also represented through the confusion matrix, which can be found in Figure 11 for the LSTM model and Figure 12 for the GRU model. The accuracy of these two models indicates that the LSTM model managed to achieve a maximum accuracy of 100%, while the GRU model achieved an accuracy of 98%.

7/7 [==============================] - 2s 163ms/step

	precision	recall	f1-score	support
0	1.00	1.00	1.00	63
1	1.00	1.00	1.00	82
2	1.00	1.00	1.00	69
accuracy			1.00	214
macro avg	1.00	1.00	1.00	214
weighted avg	1.00	1.00	1.00	214

Figure 11. LSTM evaluation performance results

7/7 [==============================] - 2s 140ms/step

	precision	recall	f1-score	support
0	0.98	1.00	0.99	63
1	0.98	0.99	0.98	82
2	0.99	0.96	0.97	69
accuracy			0.98	214
macro avg	0.98	0.98	0.98	214
weighted avg	0.98	0.98	0.98	214

Figure 12. GRU evaluation performance results

Second Validation Result

In the second validation stage, the training data were divided into 80% and the test data by 20%, with the same parameters and number of layers on the LSTM and GRU models. Although the results of this validation are not yet fully satisfactory, the accuracy ROC graph can be seen in Figures 13 and 14, where the curves generated by the LSTM model are better than those generated by the GRU model. In addition, changes in the ROC loss curve can also be observed through visualizations in Figures 15 and 16.

In addition, the results of the first validation are presented through the confusion matrix, which can be found in Figure 17 for the LSTM model and Figure 18 for the GRU model. The accuracy results of these two models show that both the LSTM and GRU models have an accuracy of 98%.

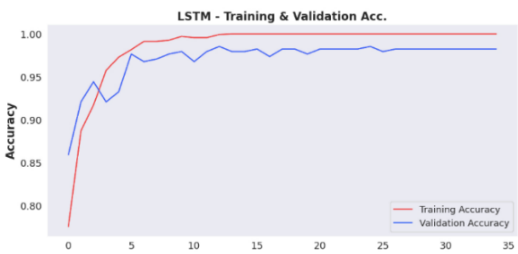

Figure 13. LSTM accuracy curve

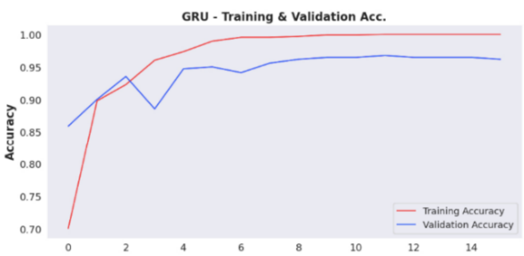

Figure 14. GRU accuracy curve

Figure 15. LSTM loss curve

Figure 16. GRU loss curve

14/14 [==============================] - 3s 168ms/step

	precision	recall	f1-score	support
0	0.98	0.97	0.98	115
1	0.99	1.00	0.99	164
2	0.98	0.97	0.98	148
accuracy			0.98	427
macro avg	0.98	0.98	0.98	427
weighted avg	0.98	0.98	0.98	427

Figure 17. LSTM evaluation performance results

14/14 [==============================] - 3s 138ms/step

	precision	recall	f1-score	support
0	0.93	1.00	0.97	115
1	1.00	0.99	0.99	164
2	0.99	0.95	0.97	148
accuracy			0.98	427
macro avg	0.97	0.98	0.98	427
weighted avg	0.98	0.98	0.98	427

Figure 18. GRU evaluation performance results

Third Validation Result

In the third validation stage, the training data were divided into 70% and the test data by 30%, with the same parameters and number of layers on the LSTM and GRU models. Although the results of this validation are still not optimal, the accuracy ROC graph can be seen in Figures 19 and 20. In addition, changes in the ROC loss curve can also be observed through visualizations in Figures 21 and 22.

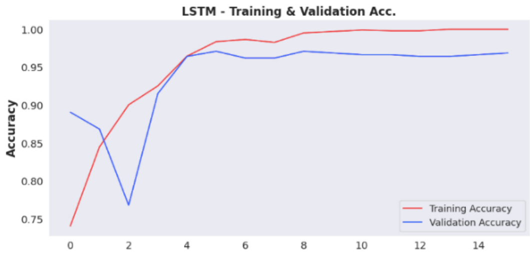

Figure 19. LSTM accuracy curve

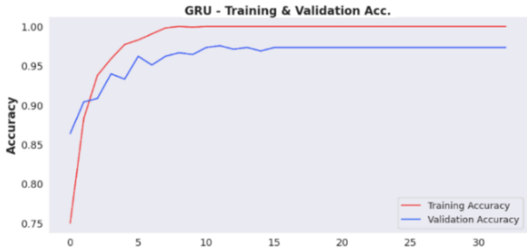

Figure 20. GRU accuracy curve

Figure 21. LSTM loss curve

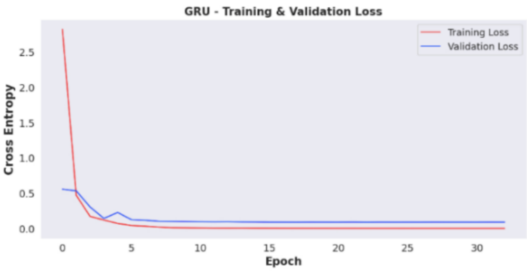

Figure 22. GRU loss curve

The evaluation results of the third version of these two models show that both LSTM and GRU have an accuracy of 97%. However, in the middle of training, the data managed to achieve a training accuracy of 100%. Furthermore, further analysis is needed to understand the factors that influence the difference in accuracy results between the two models. Apart from accuracy, details about performance metrics, such as precision, recall, and F1-score, are available in Figures 23 and 24. A more in-depth analysis of these metrics will offer a better understanding of the comparative effectiveness of both models.

20/20 [==============================] - 4s 159ms/step

	precision	recall	f1-score	support
0	0.96	0.98	0.97	190
1	1.00	0.98	0.99	231
2	0.96	0.96	0.96	219
accuracy			0.97	640
macro avg	0.97	0.98	0.97	640
weighted avg	0.98	0.97	0.98	640

Figure 23. LSTM evaluation performance results

20/20 [==============================] - 4s 142ms/step

	precision	recall	f1-score	support
0	0.91	1.00	0.95	190
1	1.00	0.98	0.99	231
2	0.98	0.92	0.95	219
accuracy			0.97	640
macro avg	0.96	0.97	0.96	640
weighted avg	0.97	0.97	0.97	640

Figure 24. GRU evaluation performance results

Fourth Validation Result

In the fourth validation stage, there was a division of training data by 60% and test data by 40%. Although the ROC accuracy of both the LSTM and GRU models reaches 96%, it can be seen that the results of this accuracy curve are still unsatisfactory, as illustrated in Figures 25 and 26.

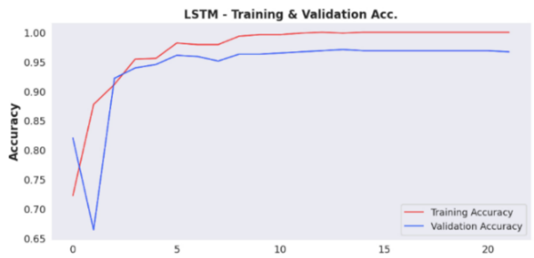

Figure 25. LSTM accuracy curve

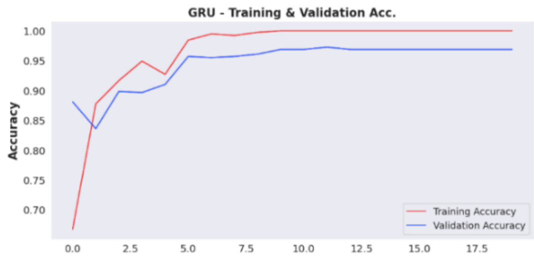

Figure 26. GRU accuracy curve

Figure 27. LSTM loss curve

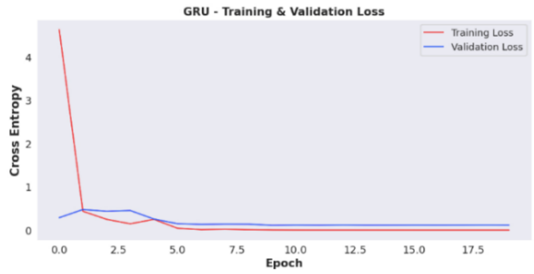

Figure 28. GRU loss curve

Meanwhile, the development of the loss curve can be found in Figures 27 and 28. With this value division resulting in lower accuracy compared to the previous three validations involving more training data, the results of evaluation performance through the confusion matrix in the form of precision, recall, and f1-scores can be seen in Figures 29 and 30. This evaluation provides further insight into the predictive and classification capabilities of LSTM and GRU models in the context of more limited datasets.

27/27 [==============================] - 5s 160ms/step

	precision	recall	f1-score	support
0	0.93	0.98	0.96	261
1	1.00	0.98	0.99	297
2	0.96	0.93	0.95	295
accuracy			0.96	853
macro avg	0.96	0.97	0.96	853
weighted avg	0.97	0.96	0.96	853

Figure 29. LSTM evaluation performance results

27/27 [==============================] - 5s 142ms/step

	precision	recall	f1-score	support
0	0.94	0.98	0.96	261
1	1.00	0.96	0.98	297
2	0.95	0.95	0.95	295
accuracy			0.96	853
macro avg	0.96	0.96	0.96	853
weighted avg	0.96	0.96	0.96	853

Figure 30. GRU evaluation performance results

Fifth Validation Result

In the fifth validation stage, the distribution of training and test data was carried out with a ratio of 50:50. From the ROC curve analysis, it can be observed that the training accuracy graphs of both methods are increasing, but the accuracy graphs at the validation stage continue to decline, resulting in unsatisfactory results, as seen in Figures 31 and 32. Similarly, there was a slight increase in the LSTM and GRU loss charts after the validation stage, as illustrated in Figures 33 and 34. This change could impact the final accuracy of both models, where LSTM reaches 96% and GRU reaches 93%.

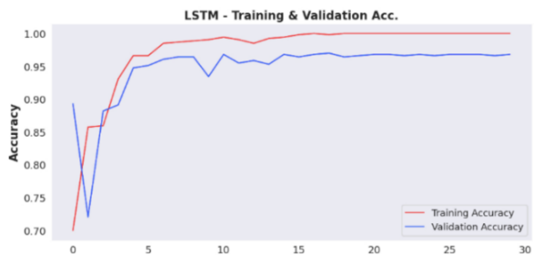

Figure 31. LSTM accuracy curve

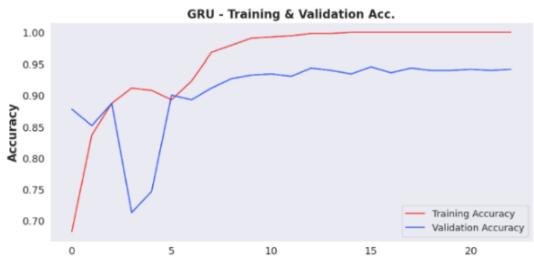

Figure 32. GRU accuracy curve

Figure 33. LSTM loss curve

Figure 34. GRU loss curve

In this fifth model, the performance comparison between LSTM and GRU models in classifying EEG emotion data can be observed through the confusion matrices depicted in Figures 35 and 36. The experimental results show that LSTM outperforms GRU with an accuracy of 96%, while GRU achieves an accuracy of around 93%. Although both models exhibited comparable precision and recall results, the significant difference in accuracy indicated the superiority of the LSTM model in this classification task.

```
34/34 [==============================] - 6s 156ms/step
              precision    recall  f1-score   support

           0       0.94      0.96      0.95       348
           1       0.99      0.98      0.98       363
           2       0.94      0.94      0.94       355

    accuracy                           0.96      1066
   macro avg       0.96      0.96      0.96      1066
weighted avg       0.96      0.96      0.96      1066
```

Figure 35. LSTM evaluation performance results

```
34/34 [==============================] - 6s 141ms/step
              precision    recall  f1-score   support

           0       0.90      0.93      0.92       348
           1       0.99      0.97      0.98       363
           2       0.90      0.89      0.90       355

    accuracy                           0.93      1066
   macro avg       0.93      0.93      0.93      1066
weighted avg       0.93      0.93      0.93      1066
```

Figure 36. GRU evaluation performance results.

Table 6. The accuracy results

Data Ratio	Accuracy	
	LSTM	GRU
90:10	100%	98.13%
80:20	98.36%	97.65%
70:30	97.50%	96.56%
60:40	96.48%	96.36%
50:50	95.87%	93.15%

The comprehensive accuracy results for LSTM and GRU, showcasing impressive accuracy rates, are meticulously presented in Table 6. This table offers a holistic perspective on the performance of both models in the EEG Emotion classification task, providing valuable insights into their comparative effectiveness. A comparison with previous studies on EEG emotion recognition is provided in Table 7, which offers a comprehensive view of how the obtained results fare about existing research findings.

Table 7. Comparison of results with previous research

	EEG Emotion	
	Method	Accuracy
Baseline	Sparse Discriminative Ensemble, 2019 [34]	77.4%
	LIBSVM classifier, 2020 [35]	82.63%
	LSTM-RASM, 2020 [36]	76.67%
	Improved Neural Network, 2020 [14]	78.92%
	SVM, 2021 [37]	84.3%
	CNN-GRU hybrid layers, 2022 [38]	97.50%
	CNN-AsMap, 2022 [39]	97.10%
	Hybrid CRNN, 2023 [40]	95.33%
	GRU-Conv, 2023 [16]	87.04%
Proposed Model	LTSM 3 layers	**100%**
	GRU 3 layers	**98.13%**

CONCLUSION

Based on the experiments conducted, both the LSTM and GRU models have been successfully implemented for emotion recognition based on EEG data. While both models performed relatively well in classifying emotions, there were variations in their performance results during various training and test data scenarios. The LSTM model consistently achieved higher accuracy compared to the GRU model, indicating a better understanding and modeling of patterns associated with human emotions based on EEG data. Although the GRU accuracy reached a good level in some validation scenarios, it remained slightly below the LSTM performance. However, it is important to note that both models showed comparable abilities in recognizing different classes of emotions, as indicated by the precision, recall, and F1-score results derived from the confusion matrix evaluation. These findings offer valuable insights into the strengths and limitations of the LSTM and GRU models in emotion recognition tasks based on EEG data.

In terms of limitations, this study acknowledges the variations in performance outcomes and suggests further exploration to enhance the accuracy of emotion recognition models. Future research directions could involve investigating more complex architectures or integrating additional features to improve the overall performance of these models in real-world applications of human emotion recognition using EEG data.

REFERENCES

[1] A. Kallipolitis, M. Galliakis, A. Menychtas, and I. Maglogiannis, "Affective analysis of patients in homecare video-assisted telemedicine using computational intelligence." Neural Computing and Applications, 2020.

[2] Z. Halim and M. Rehan, "On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning." Information Fusion 53, 2020.

[3] H. Huang *et al.*, "An EEG-Based Brain Computer Interface for Emotion Recognition and Its Application in Patients with Disorder of Consciousness." IEEE Transactions on Affective Computing, 2019.

[4] S. Ma, J. Cui, C.-L. Chen, W. Xiao, and L. Liu, "An Improved Bi-LSTM EEG Emotion Recognition Algorithm." Journal of Network Intelligence, 2022.

[5] M.-E. Wu, J.-H. Syu, and C.-M. Chen, "Kelly-Based Options Trading Strategies on Settlement Date via Supervised Learning Algorithms," *Comput. Econ.*, vol. 59, no. 4, pp. 1627–1644, 2022.

[6] S. Xu *et al.*, "Using a deep recurrent neural network with EEG signal to detect Parkinson's disease." Annals of translational medicine, 8(14), 874, 2020.

[7] J. Zhou, X. Zhang, and Z. Jiang, "Recognition of Imbalanced Epileptic EEG Signals by a Graph-Based Extreme Learning Machine," *Wirel. Commun. Mob. Comput.*, vol. 2021, p. 5871684, 2021.

[8] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG Emotion Recognition using Fusion Model of Graph Convolutional Neural Network and LSTM." Applied Soft Computing Journal, 2020.

[9] H. Yang, J. Han, and K. Min, "A Multi-Column CNN Model for Emotion Recognition from EEG Signals," *Sensors*, vol. 19, no. 21. 2019.
[10] P. Sandheep, S. Vineeth, M. Poulose, and D. P. Subha, "Performance analysis of deep learning CNN in classification of depression EEG signals," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 1339–1344.
[11] M. K. Chowdary, J. Anitha, and D. J. Hemanth, "Emotion Recognition from EEG Signals Using Recurrent Neural Networks." Electronics 11, 2022.
[12] W. Lu, Y. Wei, J. Yuan, Y. Deng, and A. Song, "Tractor Assistant Driving Control Method Based on EEG Combined With RNN-TL Deep Learning Algorithm," *IEEE Access*, vol. 8, pp. 163269–163279, 2020.
[13] X. Du *et al.*, "An Efficient LSTM Network for Emotion Recognition from Multichannel EEG Signals." IEEE, 2020.
[14] P. Nagabushanam, S. Thomas George, and S. Radha, "EEG signal classification using LSTM and improved neural network algorithms," *Soft Comput.*, vol. 24, no. 13, pp. 9981–10003, 2020.
[15] J. X. Chen, D. M. Jiang, and Y. N. Zhang, "A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification," *IEEE Access*, vol. 7, pp. 118530–118540, 2019.
[16] G. Xu, W. Guo, and Y. Wang, "Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-Conv architecture," *Med. Biol. Eng. Comput.*, vol. 61, no. 1, pp. 61–73, 2023.
[17] Y. Zhang *et al.*, "An Investigation of Deep Learning Models for EEG-based Emotion Recognition." Front. Neurosci. 14:1344., 2020.
[18] A. Romney and V. Manian, "Comparison of Frontal-Temporal Channels in Epilepsy Seizure Prediction Based on EEMD-ReliefF and DNN," *Computers*, vol. 9, no. 4. 2020.
[19] H. A. Glory, C. Vigneswaran, S. S. Jagtap, R. Shruthi, G. Hariharan, and V. S. S. Sriram, "AHW-BGOA-DNN: a novel deep learning model for epileptic seizure detection," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 6065–6093, 2021.
[20] J. J. Bird, L. J. Manso, E. P. Ribiero, A. Ekart, and D. R. Faria, "A study on mental state classification using eeg-based brain-machine interface." 9th International Conference on Intelligent Systems, IEEE, 2018.
[21] J. J. Bird, A. Ekart, C. D. Buckingham, and D. R. Faria, "Mental emotional sentiment classification with an eeg-based brain-machine interface." The International Conference on Digital Image and Signal Processing (DISP'19), Springer, 2019.
[22] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *J. Big Data*, vol. 7, no. 1, p. 28, 2020.
[23] W. K. Sari, D. P. Rini, R. F. Malik, and I. S. B. Azhar, "Multilabel Text Classification in News Articles Using Long-Term Memory with Word2Vec." Jurnal Resti, pp. 276–285, 2020.
[24] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci, "Deep Neural Networks and Tabular Data: A Survey," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–21, 2022.
[25] A. Kumar and R. Rastogi (nee Khemchandani), "Attentional Recurrent Neural Networks for Sentence Classification BT - Innovations in Infrastructure," 2019, pp. 549–559.
[26] C. S. Alfredo and D. Adytia, "Time Series Forecasting of Significant Wave Height using GRU, CNN-GRU, and LSTM." Jurnal Resti, pp. 776–781, 2022.
[27] S. Gao *et al.*, "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *J. Hydrol.*, vol. 589, p. 125188, 2020.
[28] R. Shashidhar, S. S. Tippannavar, K. B. Bhat, N. Sharma, M. Rashid, and A. Rana, "EEG-based Brain Wave Recognition using GRU and LSTM," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, 2022, pp. 1083–1087.
[29] G. Weiss, Y. Goldberg, and E. Yahav, "On the Practical Computational Power of Finite Precision RNNs for Language Recognition." arXiv preprint arXiv:1805.04908, 2018.
[30] M. Grandini, E. Bagli, and G. Visani, "Metrics For Multi-Class Classification: An Overview." ArXiv. /abs/2008.05756, 2020.
[31] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-Label Classifier Performance Evaluation with Confusion Matrix." Computer Science & Information Technology, 1, 2020.
[32] Q. H. Nguyen *et al.*, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," *Math. Probl. Eng.*, vol. 2021, p. 4832864, 2021.
[33] V. R. Joseph, "Optimal Ratio for Data Splitting." The ASA Data Science Journal, 15, pp. 531–538, 2022.
[34] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal Emotion Classification Using EEG Signal With Sparse Discriminative Ensemble," *IEEE Access*, vol. 7, pp.

40144–40153, 2019.
[35] T. Chen, S. Ju, F. Ren, M. Fan, and Y. Gu, "EEG emotion recognition model based on the LIBSVM classifier," *Measurement*, vol. 164, p. 108047, 2020.
[36] Z. Li, X. Tian, L. Shu, X. Xu, and B. Hu, "Emotion Recognition from EEG Using RASM and LSTM BT - Internet Multimedia Computing and Service," 2018, pp. 310–318.
[37] Y. Liu and G. Fu, "Emotion recognition by deeply learned multi-channel textual and EEG features," *Futur. Gener. Comput. Syst.*, vol. 119, pp. 1–6, 2021.
[38] M. Asif, M. T. Vinodbhai, S. Mishra, A. Gupta, and U. S. Tiwary, "Emotion Recognition in VAD Space During Emotional Events Using CNN-GRU Hybrid Model on EEG Signals BT - Intelligent Human Computer Interaction," 2023, pp. 75–84.
[39] M. Z. I. Ahmed, N. Sinha, S. Phadikar, and E. Ghaderpour, "Automated Feature Extraction on AsMap for Emotion Classification Using EEG." Sensors, 2022.
[40] M. Zhong, Q. Yang, Y. Liu, B. Zhen, F. Zhao, and B. Xie, "EEG emotion recognition based on TQWT-features and hybrid convolutional recurrent neural network," *Biomed. Signal Process. Control*, vol. 79, p. 104211, 2023.